

機械学習における不採用訓練事例の有効活用手法に関する研究

著者	小石 泰毅
その他のタイトル	Research on Machine Learning Method using Low Confidence Data
学位授与年度	平成30年度
学位授与番号	17104甲生工第343号
URL	http://hdl.handle.net/10228/00007208

博士学位論文

機械学習における不採用訓練事例の
有効活用手法に関する研究

生命体工学専攻 宮本研究室

学籍番号 16899007

小石 泰毅

目次

第1章 序章	1
1.1 研究の背景	1
1.2 産業における機械学習	3
1.3 限られたデータを活用する学習手法	4
1.4 本研究の目的	8
1.5 本論文の構成	10
第2章 飼いならし学習問題	11
2.1 問題設定	11
2.2 Bagging - 飼いならし学習の解法 -	13
2.3 関連する機械学習問題	16
第3章 提案手法	20
3.1 飼いならし学習問題の課題	20
3.1.1 偏りのある訓練事例集合を用いた学習	22
3.1.2 誤ったラベルを含む訓練事例集合	25
3.1.3 Baggingにおける不採用弱学習器の考察	26
3.2 提案手法のアルゴリズム	28
3.2.1 提案手法のアイデア	28
3.2.2 提案手法のアルゴリズム	30
第4章 汎用データセットを用いたシミュレーション実験	33
4.1 シミュレーション実験に用いる汎用データセット	34
4.2 実験条件	35
4.3 不採用訓練事例に対するラベル推定精度の検証	36
4.3.1 二値分類におけるラベル推定精度	36
4.3.2 多値分類におけるラベル推定精度	39
4.4 ラベル推定による野生データの推移	42
4.4.1 二値分類における野生データの推移	42
4.4.2 多値分類における野生データの推移	46

4.5	提案手法の判別能力	49
4.5.1	二値分類における判別能力	49
4.5.2	多値分類におけるラベル推定精度	57
4.5.3	提案手法の考察	64
4.6	提案手法のまとめ	69
第 5 章	ワイヤボンディングにおける抜取検査を題材とした実証実験	71
5.1	はじめに	71
5.2	機械学習のワイヤボンディング抜取検査への適用	74
5.2.1	ワイヤボンディング	74
5.2.2	ワイヤボンディングの信頼性評価とその課題	75
5.2.3	薄型 A E センサを用いた印加超音波の計測	78
5.2.4	弾性波の解析と機械学習に用いる特徴量	80
5.2.5	提案手法を用いた品質推定	81
5.3	実験	85
5.3.1	実験環境	85
5.3.2	データセットの構成	85
5.3.3	実験条件	87
5.3.4	実験結果・考察	88
5.4	結論	94
第 6 章	結論	96

第1章 序章

1.1 研究の背景

近年、第4次産業革命と呼ばれる急速な技術革新が進んでおり、情報通信技術の発展や、ビッグデータ、ロボット、AIなどの技術の進展によって社会に大きな変革を与えている。また、産業界ではCPS（Cyber Physical System）やIoT(Internet of Things)などの実社会のあらゆる情報に対してデータ化、ネットワーク化をもたらす技術が注目されてきている。これらの情報技術がもたらすデータの取得・分析・実行サイクルは、あらゆる分野における情報制約、物理制約の克服を可能としてきており、革新的なサービスの創出や供給効率性の向上の両面から産業界への応用ニーズが高まってきている。IoT、AIなどの第4次産業革命におけるコア技術はすべての産業において共通する基盤技術であり、様々な分野における事業と結びつくことで、全く新たな付加価値の充足が可能になるという期待が高まってきている。

第4次産業革命における産業構造の変革は、世界各国で重要視されており、世界の主要各国は新たな産業構造の構築を進めてきている。例えば、官民一体となって進めているドイツの“インダストリー 4.0”や、フランスの“未来の産業”，中国の“製造 2025”など、世界の主要各国はそれぞれの明確な旗を立て、第4次産業革命への対応を進めてきている。わが国でも、目指す産業の姿として“Connected Industries”というコンセプトを掲げている。“Connected Industries”とは、データを介して、機械、技術、人など様々なものが相互に連携し、新た

な付加価値の創出と社会課題の解決を目指す産業のあり方であり、モノとモノがつながるIoTのみならず、モノと人、人と人などの様々なつながりによる価値創出が実現する産業の姿を目指している [1].

第4次産業革命では、技術革新・データ利活用により、今までは対応しきれなかった「顧客の真のニーズ」への本質的な対応、新たな財・サービスの創出、グローバルな課題解決への期待が高まっている。そのため、新たな付加価値の源泉はデータにあり、取得、分析、活用のサイクルを通じて新たなニーズの創出や充足を実現する必要性が増している。また、近年のデータの利活用は、WebやSNSなどのサイバー空間から得られるバーチャルデータから、生産現場などの実環境から得られるリアルデータへと対象がシフトしている。このような状況を受け、わが国では製造業における充実した生産設備群と高い現場力に基づいた高品質なモノづくり技術を持つという強みを活かし、リアルデータの利活用による「顧客の真のニーズ」への対応や新たな価値の創出が求められている。

わが国では現在、産業におけるリアルデータの利活用が必要であるとの認識が高まる一方で、実際の利活用状況に本格的な変化はいまだ起きていない [1]. ものづくりの現場では、人手不足が顕在化する一方で、多品種少量生産などに代表されるように、顧客の高度かつ多様化したニーズに対して迅速に対応する必要性に迫られている。しかし、現状のものづくり企業の現場においては、機器間の連携が十分ではなく、限られたパターンによる製造のみ対応可能となっているため、柔軟なものがづくりが困難である場合が多い。また、少子高齢化の進展による若手技術者の減少を受け、熟練技術者が培ってきたノウハウがうまく継承されないなどの課題もある。このような課題を解決するため、Connected Industries のコンセプトのもと、情報技術を活用し様々なつながりを推進することを通じ、機器のIoT化や熟練技術の継承などを推し進め、持続的に発展可能な産業の構築を目指す科学技術が望まれている。

1.2 産業における機械学習

Connected Industries 実現のために、個々の製造現場にある高い技術力をデータとして収集し、その利活用による現場の効率化・高度化を通じた多様なニーズへの応答が重要となってくると考える。第4次産業革命におけるデータの利活用手法として最も注目を集めている技術の一つが機械学習である。機械学習は明示的にプログラミングすることなく、データや経験から学習する能力をコンピュータに与える技術であり、既存のルールベースプログラミングでは対処困難であった複雑な現象などへの応用が期待されている。機械学習を用いた推定・予測・分類を用いたデータの利活用が広く望まれており、新たな価値の創出やニーズの充足に向けて様々な分野、用途に対する基盤研究、応用研究が展開されている。実際に、製造現場における生産性向上 [2–4]、購買情報の分析による需要予測 [5–9]、異常の早期検知 [10–18]、自動走行モビリティ [19–28]、金融 [29–34]、医療 [35–39]、教育 [40, 41]、農業 [42–45]、観光 [46–48] など、様々な用途に対しての高度情報技術を活用した取組が活発に展開されている。

機械学習のアルゴリズムのうち、最も精度が高い手法として教師あり学習がある。教師あり学習は、多量のラベル付きデータから学習することで高い学習性能を示すことができる。画像処理 [49, 50] や音声認識 [51, 52]、自然言語処理 [53] などの分野で、高い判別能力を実現できる手法として近年注目を集めている深層学習も、大量のラベル付きデータから学習を行う教師あり学習に分類される。

しかし、教師あり学習によって高精度な予測・分類を行うためには、高信頼かつ大量のラベル付きデータを準備する必要がある。近年の産業は、生産ロスを抑えるための小ロット生産や顧客のニーズの多様化に応じた多品種少量生産などの生産方式への移行が進められており、今後多量のデータを収集することが難しくなってくると考えられる。また、現在生産拠点に設置されている機器はデータ取得を前提としていないものも多く、標準化もされていない。そのた

め、それぞれの機器によって取得可能なデータの種類、粒度が異なっており、各工程のデータを一元的に集約するだけでは活用可能なデータを収集することはできない。

多量のデータ収集が困難であることに加え、データの信頼性確保においても課題がある。製造業における現行の生産ラインでは目視検査などの官能検査によって品質検査を行うことも多い。官能検査では検査員の習熟度によって検査基準にばらつきが生じるため、画一的な基準による品質決定が困難である。さらに、機器やセンサにまつわるバイアスによるデータの信頼性のばらつきもあり、一つの品種に対してさえ高信頼かつ大量のデータ確保は困難な状況にある。したがって、高信頼な検査を行うことができる製品は限られてしまう。

このように、データの信頼性と量を両立するためには時間、費用、ヒューマンリソースなどの多大な収集コストがかかるため、産業におけるリアルデータの利活用における大きな障壁となっている。

これらの状況を踏まえ、リアルデータの利活用のためには、ビッグデータに代表される大量のデータの活用ではなく、むしろ少数の高信頼なデータをどのように活用するか、という課題設定が重要になってくる。したがって、リアルデータを活用し、今までにない新たな価値を創出するためには、取得可能な限られたデータを最大限活用するアルゴリズムが必要となる。

1.3 限られたデータを活用する学習手法

一般的な教師あり学習では、大量のラベル付きデータから学習を行うことで、これからテストしたい未知データに対する汎化性能を向上させることができる。しかし、前述のように高信頼かつ大量のラベル付きデータを収集することは高コストであり、産業分野への応用は大量のデータ収集が容易な事例に限られる。

少量のデータを活用する学習手法はいくつか提案されている。本節では、そ

れらについて概説する。

半教師あり学習

半教師あり学習は、高信頼なラベルが与えられているラベル付きデータと、ラベルの与えられていないラベルなしデータを併用する学習問題である。半教師あり学習は、ラベルなしデータに対して与えられるべきラベルを推定し、疑似的なラベル付きデータとすることで、教師あり学習と同様に学習を行う手法である。

半教師あり学習の応用例として、意味獲得や単語分割などの自然言語処理 [54–57] や画像処理における一般物体認識 [58, 59] が挙げられる。

半教師あり学習では、ラベル付きデータから推測できるデータの分布などを用いてラベルなしデータにラベルを与える。したがって、ラベルなしデータに対する分布の推定精度、与えたラベルの確からしさが重要となる。一般に、少量のラベル付きデータからの分布の推定は困難であるため、データ生成時の暗黙知などを考慮した分布推定による補強が必要となる。

能動学習

少ないデータから学習効率を上げる機械学習として、能動学習がある。能動学習はラベル付きデータを作成する際に最大の効果を発揮するように、ラベルを与えるデータを選択する学習手法である。一般的に多量のデータに正しくラベルを与えることは高コストであるため、精度の高い学習器を生成するためにより効果の高いデータに対して選択的にラベル付けを行うことで、ラベル付きデータの収集コストを低減させている。

能動学習の応用例として、文脈理解などの自然言語処理 [60–63] や、製造現場などにおけるシステムの異常検知 [64–66]、システムのパラメータ設計 [67, 68]、

バイオインフォマティクス [69] などが挙げられる。

能動学習は、ラベルに専門家や熟練者の知識が必要な場合や、創薬などラベル付けに長い時間が必要な状況など、ラベル付けが時間・ヒューマンリソースの面で高コストであり、ラベル付きデータの大量収集が困難である場合に有効な学習手法である。一方で、ラベル付けが容易にできる環境がないと有効に作用しないため、高い専門性を持つ人物によるラベル付けが必要であるという点で、能動学習は収集コストが高くなりやすい手法であるといえる。

転移学習

転移学習は二種類のラベル付きデータを扱う学習問題である。転移学習ではターゲットドメインとソースドメインと呼ばれる二種類のデータを扱う。ターゲットドメインはこれから学習したい問題であり、ソースドメインはターゲットドメインに類似した別の問題である。転移学習では、あらかじめ多量に収集し、管理されているソースドメインからターゲットドメインの予測・分類に有効な知識や情報を学習し、ターゲットドメインに対して知識を転移することで、ターゲットドメインの収集コストを抑制する学習手法である。

転移学習の応用例として、医療分野 [70–72]、バイオ・マテリアルズインフォマティクス [73–77]、行動認識 [78]、画像処理 [79]、製造効率化 [80] などが挙げられる。

転移学習では、ソースドメインから転移させる知識や情報の選択が重要となり、誤った知識の転移は「負の転移 (negative transfer)」[81] を引き起こし、かえって学習精度を悪化させてしまう。また、ソースドメインには正しくラベルがつけられている必要があるため、データ収集に係るコストを削減するためには扱うドメインに関する専門的な知見が必要となる。

上記の手法は、いずれも高信頼な少数のデータからの学習を補完するために、

別のデータセットを活用する手法である．半教師あり学習では，ラベルなしデータに対するラベル推定が必須となるが，生産現場などでは推定するための情報として活用できる高信頼なデータは少なく，多くの場合でラベル推定の精度は十分ではない．また，能動学習や転移学習では，活用可能なデータセットはこれから解きたい問題に類似したデータであり，そのようなデータが容易に準備できることはまれである．さらに，これらの手法は学習の補完に用いるデータセットは高信頼なラベルが与えられる必要がある．製造業などの生産現場では，むしろ誤ったラベル付けをされたデータを含む低信頼なデータが手に入りやすい．そのため，生産現場への適用のためには，低信頼なデータセットを用いて高信頼なデータを補完する手法が望ましい．

1.4 本研究の目的

第4次産業革命における産業構造の変革に伴い、生産現場では多品種少量生産などによる多様化する消費者ニーズへの対応と、生産性・設備稼働率の向上は必須となってきている。これらの変動への対応と効率性を両立し、高付加価値な製品を市場に流通させるためには、従来の人海戦術による改善活動ではすでに限界を迎えており、機械学習などの高度情報処理技術を活用した新たな生産体制が求められている。多様なニーズを充足可能な新しい生産体制を迅速に構築するうえで重要なことは、全体最適化を行うための改善活動サイクルを早めることである。既存の機械学習手法の多くは、学習目標に従って正確にラベル付けされた多量のラベル付きデータから学習することで高い学習能力を獲得しているが、産業における高信頼なデータの収集はコストが高いため、時間的・費用的制約が大きくなってしまう。さらに、製造現場では各拠点ごとの検査員による官能検査や機器バイアスなどがあるため、生産拠点から一元的に集約したデータセットをそのまま扱うことも難しい。したがって、高信頼かつ多量のデータを収集することは困難であり、データ収集に時間、費用、ヒューマンリソースの面で多大なコストを要し、それらによる全体最適化のサイクルの遅延が新たな生産体制構築における障壁となっている。そのため、大量のラベル付きデータを要する機械学習アルゴリズムよりも、むしろ誤りを含んだ限られたデータを最大限活用する機械学習アルゴリズムが求められており、そのような手法を採用することで、既存の生産体制でボトルネックとなっていたデータ収集コストを低くすることが望ましい。

誤ったラベルをもつデータを扱う機械学習問題として、神宮らが提案した飼いならし学習問題 [82] が提案されている。飼いならし学習では、“野生データ”と呼ばれる誤ったデータを含む低信頼なデータセットから学習に用いるデータを選択的に利用することで、これから学習したい問題である“飼育データ”に対

する学習精度を向上させる機械学習問題である．飼いならし学習問題は，同一の事例を扱いつつ，ラベルの水準がそれぞれ異なるという転移学習の特殊な例であるといえる．

しかし，飼いならし学習問題では，飼育データと同水準でラベルを与えられた野生データを選択的に活用するにとどまっている．前述のように，一般的には野生データにどれだけの飼育データと同水準でラベルを与えられたデータが含まれているかは未知であり，飼いならし学習の手法ではどれだけのデータが活用できるかは未知となる．そのため，各生産現場からデータを集約したとしても，活用可能なデータ数が増加するとは限らない．また，各生産現場から取得可能なデータ数も限られているため，迅速かつ柔軟な新しい生産体制の構築のためには，一元的に集められた限られたデータを最大限活用することが重要となる．

そこで本研究では，飼いならし学習問題を題材として，収集可能な限られたデータを最大限活用する手法を提案する．具体的なアルゴリズムとして，飼いならし学習問題における野生データのうち，所望する基準とは異なる基準でラベルを与えられたデータに対して，飼育データと同等の水準によるラベル修正を施す．修正した野生データを用いて再度学習することで，データの利活用を促す新しいアルゴリズムを提案する．提案したアルゴリズムに対して，汎用データセットを用いた実験を通じて妥当性を検討し，従来手法との精度比較を行う．また，本研究では製造業を対象としているため，半導体接合技術であるワイヤボンディングの信頼性検査を題材とした実証実験を行う．

1.5 本論文の構成

本論文は6章構成となっている。以下の手順で研究成果を報告する。第1章では、研究の背景と提案手法の目的を述べる。第2章では、飼いならし学習問題に関する定義とその解法である BaggTaming について説明する。また、少数のデータから学習する手法を紹介し、飼いならし学習問題と比較することで、飼いならし学習問題の位置づけを明確にする。第3章では、飼いならし学習問題に基づく新しい手法の提案と、そのコンセプトについて述べる。第4章では、汎用データセットを用いた実験を通じて、提案手法の妥当性を検証する。また、従来手法との精度比較を行うことで、提案手法の評価を行う。第5章では、半導体接合技術であるワイヤボンディングを題材とした実証実験により、実環境における提案手法の有効性を検討する。最後に、第6章で本研究の提案手法のまとめと新たな知見、想定される波及効果と提案手法の今後の展開についてまとめる。

第2章 飼いならし学習問題

信頼性の高い訓練事例集合と信頼性の低い訓練事例の二種類の訓練事例を扱う学習問題として、飼いならし学習問題 [82] が提案されている。本章では、飼いならし学習問題の扱う問題設定と提案されている解法について解説する。また、飼いならし学習問題に関連する機械学習問題を紹介し、それらとの比較を通じて飼いならし学習問題の扱うべき問題を明確にする。

2.1 問題設定

飼いならし学習問題では、以下の二種類の訓練事例集合を扱う。

- 飼育データ：少数の信頼性の高い訓練事例集合
- 野生データ：多数の信頼性の低い訓練事例集合

飼育データと野生データに含まれる訓練事例は、すべてラベルが与えられている。飼育データに与えられたすべてのラベルは厳密に管理されており、これから学習したい学習目標と一致している。しかし、一般に管理コストの高い訓練事例集合は収集コストも高くなり、多数の飼育データを確保することは困難である。したがって、飼育データに含まれる訓練事例数は少なくなるため、飼育データのみを用いた学習では高い学習性能（高精度な予測性能、未知データに対する汎化性）を得ることはできない。一方、野生データのラベルは厳密には管理されておらず、学習目標に従って正しくラベルを与えられた訓練事例と、学習目

標とは異なる基準でラベルを与えられた訓練事例が混在している。したがって、野生データに含まれる訓練事例は完全には信頼することができない。また、野生データは管理・収集コストが低いため、多数の訓練事例を確保することは容易であるが、どの訓練事例が学習目標に従ったラベルを与えられているかは未知であり、いくつ含まれているかも未知である。そのため、野生データのみを用いた学習では、高い学習性能を得ることは困難となる。飼いならし学習問題の目標は、飼育データと野生データを相互補完的に利用することで、飼育データのみを用いた学習よりも高い学習性能を得ることである。

飼いならし学習問題の具体例について述べる。ここでは、クラス分類問題について扱う。対象を特徴ベクトル x で表し、その対象が分類されるクラスを c と記す。特徴ベクトルとクラスの対が訓練事例となり、 (x, c) と記述する。訓練事例集合 (x_i, c_i) から、任意の特徴ベクトル x に対して、分類されるべき適切なクラスを予測できる学習器を得ることがクラス分類問題である。

通常のクラス分類問題では、一種類の訓練事例集合を扱う。すべての訓練事例は、獲得すべき学習目標を表す同時分布 $P[x, c]$ から独立に生成されているとみなされる。一方、飼いならし学習問題では、飼育データと野生データの二種類の訓練事例集合を扱う。飼育データに含まれる訓練事例は、通常のクラス分類問題と同様に、獲得すべき学習目標を表す同時分布 $P[x, c]$ から独立に生成されているとみなす。この飼育データを $D_T = \{(c_i, x_i)\}_{i=1}^{N_T}$ と表す。ここで、 $N_T = |D_T|$ である。一方、野生データには、 $P[x, c]$ から独立に生成される訓練事例と、別の概念を表す分布 $P_N[x, c]$ から独立に生成される訓練事例が混在している。また、どの事例が $P[x, c]$ から生成された訓練事例であるかは未知である。この野生データを $D_W = \{(c_i, x_i)\}_{i=1}^{N_W}$ と表す。ここで、 $N_W = |D_W|$ である。野生データは飼育データよりも十分に多いため、 $N_W \gg N_T$ となる状況を仮定する。

2.2 Bagging - 飼いならし学習の解法 -

飼いならし学習問題における手法として、Baggingが提案されている [82]. Bagging は、アンサンブル学習の一種であるバギング [83] をベースとした手法である. バギングの一般的な手続きを以下に示す.

1. for t in $1 : T$
2. 訓練事例集合 D から訓練事例を復元抽出し, 訓練事例集合 D_t を得る.
3. D_t を用いて弱学習器 f_t を生成する.
4. end for
5. 得られた弱学習器 $f_1 \cdots f_T$ の凝集により, 最終結果を得る.

このバギングで予測精度が向上する理由は, バイアス-バリエーション理論によって簡潔に説明されている [84]. バイアス-バリエーション理論では, 学習の汎化誤差は予測に用いたモデルに由来する誤差 (バイアス), 学習に用いた訓練事例に由来する誤差である (バリエーション), 本質的に減らせない誤差 (ノイズ) の三つの成分に分解できる. 多様な関数を近似可能な低バイアスなモデルを用いることでバイアスを抑制することができるが, 訓練事例のサンプリングの揺らぎに由来するバリエーションは大きくなる. また, 高バイアスなモデルでは関数を近似する能力が低く, バイアスが高くなるが, サンプリングの揺らぎにロバストであるためバリエーションは小さくなる. したがって, バイアスとバリエーションはトレードオフである. バギングは, 様々な訓練事例集合を生成し, それらから学習した弱学習器を凝集することで, バイアスを一定に保ちつつサンプリングに由来するバリエーションを抑制することができる. そのため, 低バイアスなモデルを採用することで, バリエーションのみを積極的に抑制することができるため, 汎化誤差を小さくできる.

BaggTaming は、このバギングの考えを活用した手法である。バギングでは、与えられた訓練事例集合から復元抽出を行い、バリエーションの抑制を図る。さらに、訓練事例集合から得られる特徴ベクトルに乱数を加え、より多様な訓練事例集合から弱学習器を生成することで、バギングの性能を向上できることが報告されている [84]。BaggTaming では、飼育データよりも多様な訓練事例集合である野生データから復元抽出を行うことで、バリエーションの積極的な抑制を図っている。ただし、野生データは学習目標とは異なる分布から生成された訓練事例を多数含み、復元抽出によってそれがいくつ含まれるかは未知となる。そのため、すべての弱学習器を用いて最終結果を算出した場合、学習性能が低くなってしまう恐れがある。そこで、学習目標に基づいて生成されている飼育データを利用し、生成した弱学習器に対して採用判定を行う。生成した弱学習器を用いて、飼育データのラベルを推定し、正答率を算出する。弱学習器の正答率が十分に高ければ、その弱学習器の学習に使用された訓練事例集合には多数の目標概念に従う訓練事例が含まれているとみなし、その弱学習器を採用する。弱学習器の飼育データに対する正答率が十分でない場合、訓練事例集合には目標概念とは異なる分布から生成された訓練事例が多く含まれるとみなし、その弱学習器は廃棄する。採用された弱学習器を用いて凝集することで、最終結果を算出している。以上の BaggTaming の手続きを図 2.1 に示す。

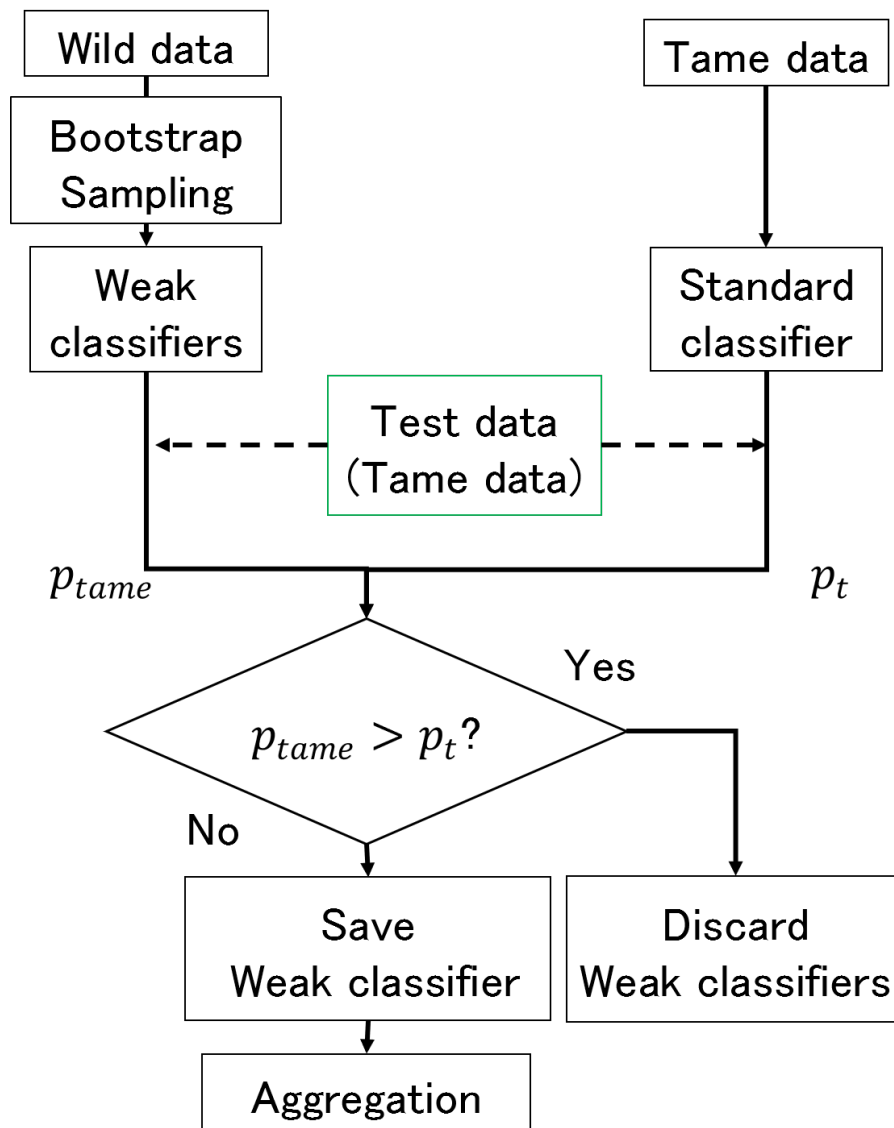


図 2.1: Algorithm of BaggingTaming

2.3 関連する機械学習問題

飼いならし学習と同様に，二種類の訓練事例集合を扱う学習問題がいくつか提案されている。

半教師あり学習は，ラベルを与えられているラベルあり訓練事例集合と，ラベルの与えられていないラベルなし訓練事例集合を併用する学習問題である。半教師あり学習の代表的な手法として，self-training [85]，Co-training [86]，Graph-based SSL [87] などがある。self-training は，分類器 f を用いてラベルなし訓練事例集合のラベルを推定する。ラベル推定した訓練事例から確信度を計算し，高い確信度で推定できた訓練事例をラベルあり訓練事例とみなして再度学習を行う。これを繰り返すことで，分類器の性能向上を図っている。Co-training は，訓練事例が二つ以上の素性に分類できる場合に有効な学習手法である。Co-training では，それぞれの素性に対して分類器を個別に使用する。説明のために，2種類の素性に分類可能な事例について，分類器 f_1 と分類器 f_2 を用いる場合を想定する。まず，ラベル付きデータのみを用いて分類器 f_1 ，分類器 f_2 をそれぞれ個別に学習する。次に，学習した分類器 f_1 ，分類器 f_2 を用いて，ラベルなし訓練事例についてラベルの推定を行う。分類器 f_1 を用いた推定の結果，高い確信度でラベル推定を行えたラベルなし訓練事例は，次回の分類器 f_2 での学習に用いる。同様に，分類器 f_2 を用いた推定によって高い確信度でラベル推定を行えたラベルなし訓練事例を，次回の分類器 f_1 での学習に使用する。これを繰り返すことで，分類器の性能向上を図っている。self-training は単一の学習器を用いてラベルなし訓練事例の推定を行うが，Co-training では複数の分類器を用いてラベル推定をしている。さらに，ある学習器の次回の学習に用いるラベルなし訓練事例は，別の分類器の各進度によって決定されている。Graph-based SSL はラベルあり訓練事例とラベルなし訓練事例を区別せず，同時に扱う手法である。構築したグラフ上でラベルあり訓練事例からラベルなし訓練事例の類似度を計算

し、ラベルを伝搬させることで、ラベルなし訓練事例のラベルを推定する。例えば、ラベルなし訓練事例の近傍データの近くにラベル 1 のラベルあり訓練事例が多く有れば、ラベル 1 に分類される。ラベルなし訓練事例のラベルを推定し、通常のラベルあり訓練事例集合とみなして学習を行う手法である。

self-training と Co-training はラベルなし訓練事例に対する確信度でラベルを更新するが、仮定された訓練事例は正しく推定されている必要がある。しかし、確信度の基準を高くした場合は多くのラベルなし訓練事例を活用できず、低い場合は学習全体の性能低下につながる。さらに、Co-training ではそれぞれの分類器は条件付独立であり、個々の分類器による学習精度も高い必要がある。また、Graph-based SSL は異なるラベルを与えるべき事例に対して、類似度が大きくなるような事例に対して適用できない。

半教師あり学習は、ラベルなし訓練事例を扱うという点で飼いならし学習問題と決定的に異なる。半教師あり学習では、少数のラベルあり訓練事例から分布を推定するなど多くの制約を持つが、飼いならし学習は多数の訓練事例から分類器を生成するため、分布の推定に係る精度の揺らぎがないという点でも異なっている。

転移学習は二種類のラベル付き訓練事例集合を扱う学習問題である。転移学習ではこれから学習したい目標となるターゲットドメインと、それに類似した問題であるソースドメインの二種類の訓練事例集合を扱う。ターゲットドメインとソースドメインの間にある違いは想定する問題によってさまざまであるが、多くの場合はターゲットドメインとソースドメインの分布が異なっている状況を扱う。ターゲットドメインとソースドメインの両方にラベルが与えられている転移学習問題は、帰納転移学習と呼ばれている [88]。帰納転移学習では、ほとんどの場合で分布の違いを扱っている。すなわち、ターゲットドメインとソースドメインの特徴ベクトルとラベルの定義域が同じであるが、同時確率が異なる

ため、うまく一致させるように学習を行う。また、少ない例ではあるが、特徴ベクトルの定義域が異なる、ラベルの定義域が異なるなどのケースもある。また、ソースドメインにラベルが与えられていない転移学習問題もあり、多くの場合は既存の特徴構築問題と同義である。

帰納転移学習は、二種類のドメインを扱うという点で飼いならし学習と似ているが、ターゲットドメインとソースドメインの両ドメインに与えられたラベルは、それぞれのドメインにおいて信頼できるという点が異なる。飼いならし学習問題は、二種類のドメインにそれぞれラベルが与えられているが、ソースドメインのラベルは完全には信頼できない問題であり、また同時確率や分布も異なっている。二種類のラベル付きドメインを扱う学習問題という点で、飼いならし学習問題は帰納転移学習の特殊な例であるともいえる。

また、少ないデータから学習効率を上げる機械学習として、能動学習がある。能動学習は、既存のデータベースから学習した後、学習に有用な訓練事例に対してラベル付けを要求し、それらを考慮して再度学習を行う手法である。能動学習には、プールベース能動学習 [89]、ストリームベース能動学習 [90]、クエリ生成型能動学習 [91] などがある。プールベース能動学習は、大量のラベルなし訓練事例から学習に有効な事例を選択し、ラベルを与えてもらうことで学習に活用する手法である。ラベル付けの効果が高い訓練事例を選択できるため、ラベル付けコストが高い事例などへの応用が期待されている。ストリームベース能動学習は、順番にやってくる訓練事例に対して、ラベル付けを要求するかどうかを決定する学習法である。訓練事例の入手コストが極端に低い場合などに有効な手法とされる。クエリ生成型能動学習は、入力空間内のあらゆるラベルなし訓練事例のラベルを問い合わせる。ここで問い合わせる訓練事例は実際にサンプリングされるもの以外も含まれており、実際のサンプリングによって得られる訓練事例よりも多様な事例に対する学習が可能となる。

能動学習は、多数の訓練事例から学習に有効な訓練事例を活用するという点で飼いならし学習と似ているが、選択された訓練事例に対して人手によるアノテーションを要求する部分が明確に異なる。

以上のことから、飼いならし学習問題は他の学習手法と比較して、二つのラベル付き訓練事例を用いるが、一方の訓練事例に与えられたラベルが誤りを含み、完全には信頼できないデータセットを扱うという点が異なるといえる。また、飼いならし学習は二種類のラベル付き訓練事例を扱い、多量の訓練事例集合から学習した知識を少量の訓練事例集合に転移するという点で帰納転移学習に似ている。したがって、飼いならし学習問題は転移学習の特殊な例であると捉えることもできる。

第3章 提案手法

本章では，飼いならし学習問題における従来手法の課題と，その課題の克服を目指す提案手法について議論する．

3.1 飼いならし学習問題の課題

飼いならし学習問題は，信頼性の高い教師ラベルを持つ少数の飼育データと，信頼性の低い教師ラベルを持つ多数の野生データを扱う学習問題である．飼いならし学習の従来手法である BaggingTaming では，野生データからランダムサンプリングによって複数の弱学習器を作成し，飼育データに対する正判別率を用いてフィルタリングを行うことで，弱学習器の採用・不採用を判定している．採用された弱学習器のみを用いて凝集を行うことで，飼育データのみから学習を行った学習器よりも高い汎化性能の獲得を目指している．したがって，野生データから飼育データと同水準の教師ラベルを持つ訓練事例を選択的に活用している．しかし，飼いならし学習では野生データに含まれる訓練事例のうち，いくつかの訓練事例が活用可能であるかは通常未知である．つまり，いくつかの訓練事例が同時分布 $P[x, c]$ から生成されているか，どの訓練事例が $P[x, c]$ から生成されているかは未知となっている． N_W のうち，多くの訓練事例が $P_N[x, c]$ から生成されている場合， $P[x, c]$ 由来の訓練事例を多く含む学習器がうまく生成できず，弱学習器の採用が困難になるという問題がある．弱学習器の採用数が少ない場合，多様な訓練事例集合を含むという野生データの利点を最大限活用でき

ず, BaggTaming の性能向上は限定的になってしまう. 本節では, BaggTaming が有効に作用しない状況について具体例を用いて論じる.

BaggTaming では, 野生データからランダムサンプリングを用いて弱学習器を生成し, 採用判定によって選択的に利用している. したがって, 多くの弱学習器を採用するほど, 野生データを有効に活用していると考えられる. この BaggTaming の採用判定において, 不採用となった弱学習器に注目する. 不採用となった弱学習器には, 以下の状況のいずれかを満たす訓練事例集合から学習していると考えられる.

1. サンプリングした訓練事例の分布に偏りのある訓練事例集合
2. $P_N[x, c]$ に由来する訓練事例を多く含む訓練事例集合

まず, これらの状況について説明するため, 入力特徴量 2 次元 (x_1, x_2) , 二値ラベル (label1, label2) を持つ疑似データを用いた簡易な例を示す. まず, 図 3.1 に示すように飼育データが分布している場合について考える. 飼育データのラベル L_i は以下の式に基づいて与えられている.

$$L_i = \begin{cases} 0 & (x_1 x_2 \geq 0) \\ 1 & (otherwise) \end{cases}$$

つまり, x_1 と x_2 のなす二次元特徴量空間において, 第一象限と第三象限に存在する訓練事例には label0 を与え, 第二象限と第四象限に存在する訓練事例には label1 を与えている. したがって, この疑似データの判別境界は図 3.2 に示す点線のように得ることを期待している.

しかし, 実際にこの疑似データから学習した判別境界は図 3.3 のようになる. これは, 学習に用いる訓練事例の数が少ないことに起因すると考えられる. 通常, アンサンブル学習などに用いる低バイアスな学習器は, 学習に用いる訓練事例の分布によって判別境界を決定する. そのため, 学習用の訓練事例集合に

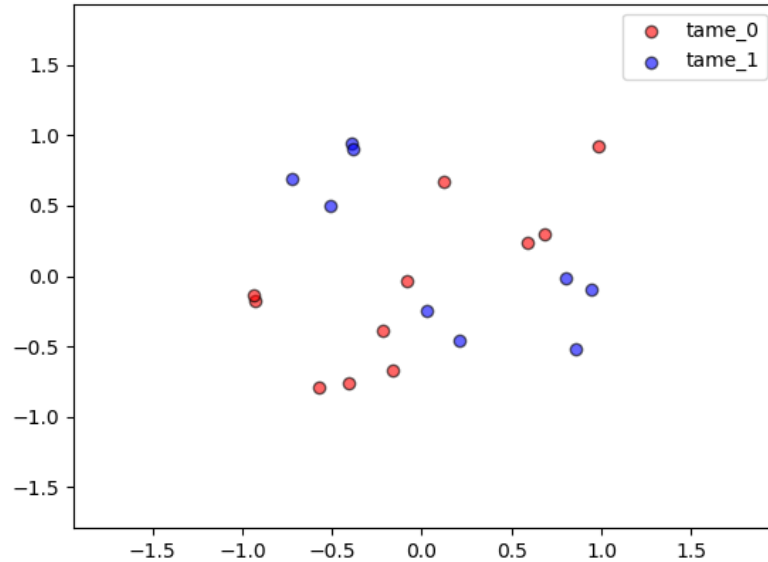


図 3.1: Pseudo tame data for explanation

意図しない偏りなどがある場合、その偏りを強く反映した判別境界を生成する。一般に学習に用いる訓練事例が少なくなるほど、この偏りは大きくなるため、所望する判別境界を得ることが困難となる。

3.1.1 偏りのある訓練事例集合を用いた学習

復元抽出によってサンプリングされた訓練事例集合に偏りがあった場合について考察する。訓練事例集合の偏りは、主に以下の二つの状況に分離することができる。

- 野生データが飼育データの一部の近傍に分布する場合
- 野生データが飼育データを包囲するように分布する場合

まず、野生データから復元抽出した訓練事例が、飼育データ近傍にまとまって分布する場合について考察する。図 3.4 は一方のラベルを持つ訓練事例が特

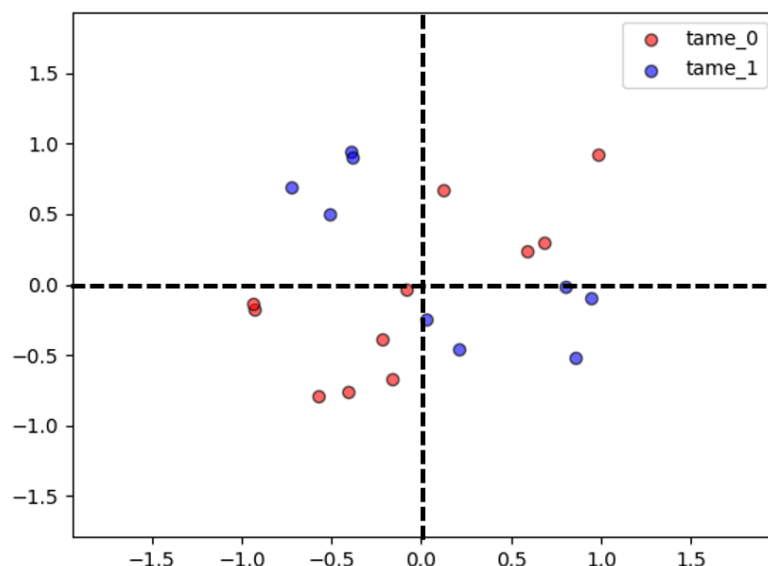


図 3.2: Desired discrimination boundary

微量空間上でまとまって存在する場合を表している．図 3.4 の場合，label1 を与えられた訓練事例が第二象限に集中的に分布しており，第四象限にはまったく存在していない．そのため，判別境界は第二象限とその他を分離するように決定されている．このような判別境界は，第四象限にある飼育データを適切に分類することができず，飼育データに対する判別能力は低くなってしまう．図 3.4 は label1 の野生データがすべて第二象限に分布しているが，いくつかの訓練事例が第四象限に分布する場合でも同様の判別境界が得られると考えられる．このことからわかるように，このような分布の偏りを持つ訓練事例集合から学習した弱学習器は飼育データに対する判別能力が低くなるため，BaggTaming の手続きにおいて採用されにくくなる．

次に，野生データから復元抽出した訓練事例が，飼育データを包囲するように分布した場合について考察する．図 3.5 はサンプリングした訓練事例が飼育

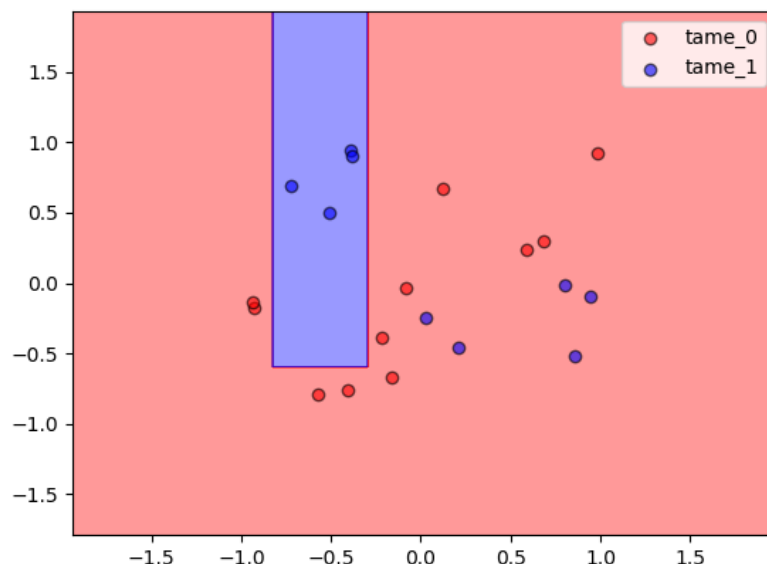


図 3.3: Discrimination boundary when tame data is used

データを包囲するように分布した場合を表している．野生データは飼育データよりも多様な訓練事例集合であるため，対象とする問題のばらつきが大きい場合はこのような分布が多く存在すると考えられる．図 3.5 の場合，野生データからサンプリングされた訓練事例は飼育データよりも広く分布している．そのため，隣り合うクラス間のどこに判別境界が引かれるかは未知となり，通常はどちらかのクラスに近接するような判別境界となる．そのため，所望する判別境界とは異なる判別境界を学習してしまい，結果的に飼育データに対する判別能力が低くなる．サンプリングした訓練事例が飼育データから大きく離れている場合，判別境界が決定可能なマージンが大きくなるため，所望する判別境界を得られる確率は低くなると考えられる．そのため，このような場合も飼育データ近傍に判別境界が存在しなくなる可能性が高いため，飼育データに対する判別能力は低くなり，BaggTaming の手続きにおいて採用されにくくなる．

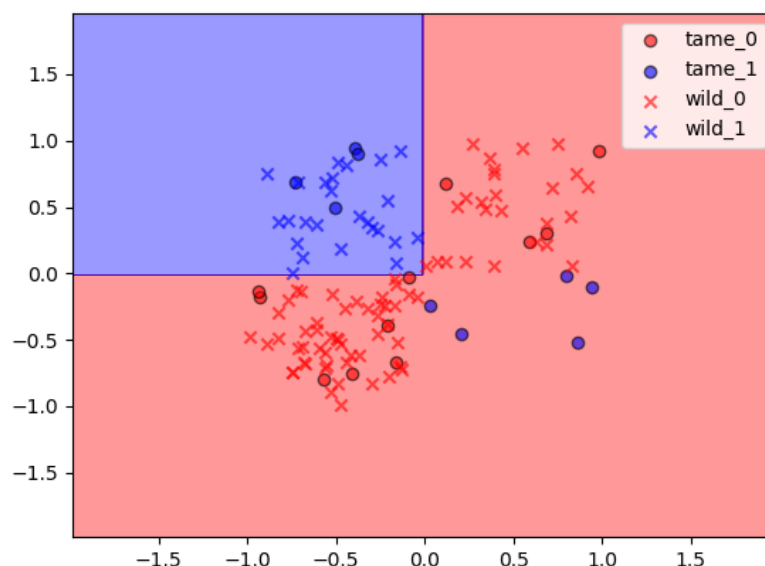


図 3.4: Discrimination boundary for biased data in the vicinity of tame data

3.1.2 誤ったラベルを含む訓練事例集合

サンプリングされた訓練事例集合に、 $P_N[x, c]$ に由来する訓練事例が多く含まれている場合について考察する。 $P_N[x, c]$ に由来する訓練事例、は $P[x, c]$ から由来される訓練事例とは異なった分布から生成されている。 飼いならし学習問題において、 $P_N[x, c]$ に由来する訓練事例はラベルの付与規則が飼育データと異なる場合が多い。 つまり、 $P_N[x, c]$ に由来する訓練事例は所望する基準とは異なる基準でラベル付けをされている訓練事例である。 このような訓練事例を多く含む訓練事例集合から学習する場合、 所望する $P[x, c]$ の概念をうまく獲得することができない。

図 3.6 に、多くの $P_N[x, c]$ から生成された訓練事例を含む訓練事例集合からの学習例を示す。 図 3.6 で表す野生データの訓練事例に与えられたラベル L_i は、以

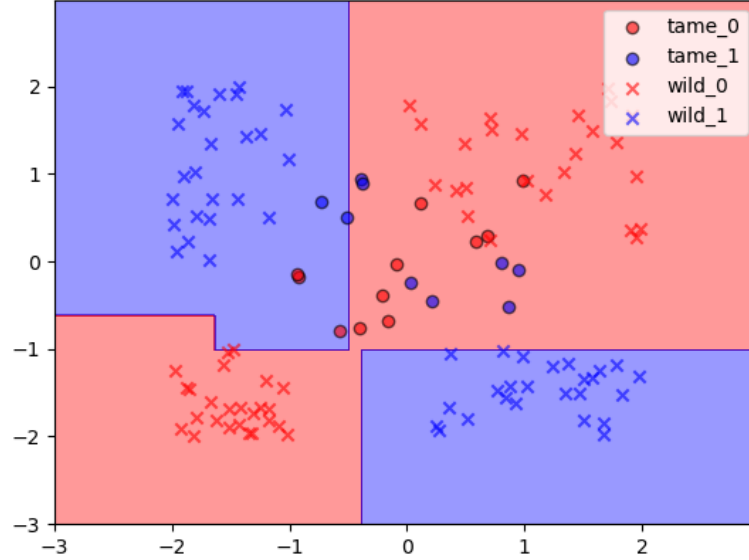


図 3.5: Discrimination boundary for data that is biased towards tame data

下の式に基づいて与えられている．

$$L_i = \begin{cases} 0 & (x_1 \geq 0) \\ 1 & (otherwise) \end{cases}$$

図3.6からわかるように，サンプリングされた訓練事例から学習された判別境界は，飼育データに対する判別能力が低くなってしまう．したがって，このような弱学習器はBaggTamingの手続きにおいて採用されにくい．

3.1.3 BaggTamingにおける不採用弱学習器の考察

上述の事例では，野生データから復元抽出した訓練事例集合から学習した弱学習器が，学習目標に対して低い判別能力を示す例を示している．3.1.1ではサンプリングされた訓練事例の分布の偏りに起因して弱学習器の判別能力が低下することを示している．また，3.1.2ではサンプリングされた訓練事例に，所望

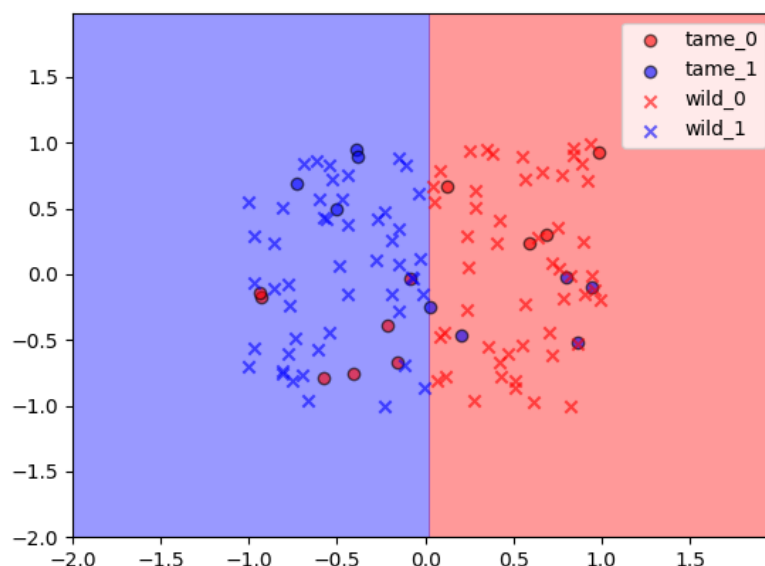


図 3.6: Discrimination boundary when erroneous data is used

する基準とは異なる分布である $P_N[x, c]$ に由来する訓練事例が多く含まれることに起因する弱学習器の判別能力の低下を示している。ここで重要なことは、3.1.1 で示した判別能力の低い弱学習器は、すべて $P[x, c]$ に由来する訓練事例であることに對し、3.1.2 で示した弱学習器は $P_N[x, c]$ に由来する訓練事例が含まれているということである。つまり、従来手法で不採用となる弱学習器には、学習目標に従ってラベルを与えられた事例とそうではない事例が混在している。

BaggTaming では、復元抽出を用いて複数の訓練事例集合を生成することにより、3.1.1 で示した弱学習器が生成されていても、別の弱学習器でその訓練事例を扱うことが可能となっているが、 $P_N[x, c]$ に由来する訓練事例は判別能力の低下を促すため、そのような訓練事例を多く活用することはできない。一般に、野生データに $P_N[x, c]$ に由来した訓練事例がいくつ含まれているかは未知である。実環境での運用を想定した場合、意図せずに所望する基準とは異なる訓練

事例 ($P_N[x, c]$ に由来した訓練事例) を多く含む可能性があり, そのような場合 BagTaming では野生データを有効に活用できず, 大幅な性能向上は期待できない. したがって, 多くの状況に対応するためには, 野生データの中に $P_N[x, c]$ に由来する訓練事例がいくつ含まれていても性能向上可能なアルゴリズムが望ましい.

そこで, 本論文では野生データを最大限活用した学習を行うために, $P_N[x, c]$ に由来する訓練事例を抽出し, それらのラベルを飼育データと同じ学習目標で再ラベリングするアルゴリズムを提案し, 従来手法よりも多くの野生データを活用し, 様々な状況におけるさらなる性能向上を目指す.

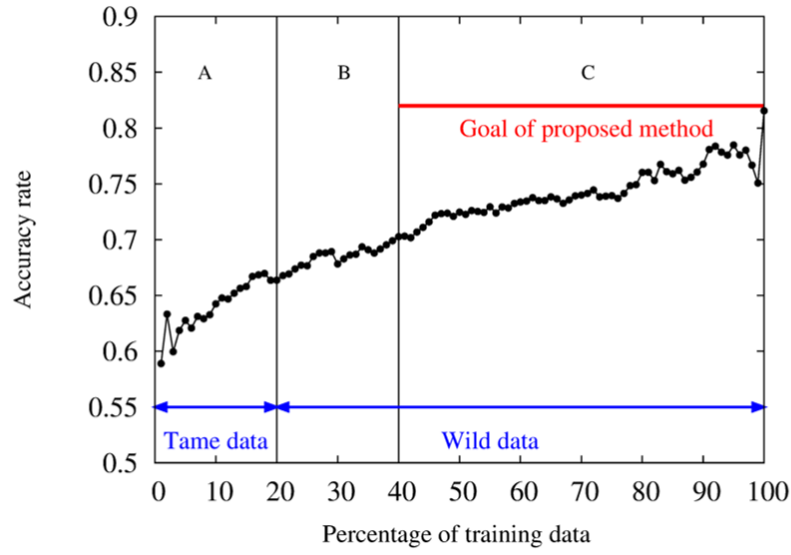
3.2 提案手法のアルゴリズム

3.2.1 提案手法のアイデア

提案手法は, 飼いならし学習問題における解法である BagTaming を拡張し, 野生データの利活用を促す手法を提案する. 飼いならし学習問題は, 飼育データと野生データを併用した学習問題であり, 前述のように野生データには学習したい目標概念である $P[x, c]$ に由来する訓練事例と, 学習したい目標概念とは異なる概念に従う $P_N[x, c]$ に由来する訓練事例が混在している. 神寫らが提案した飼いならし学習の解決法である BagTaming は, 野生データからブートストラップサンプリングを用いて複数の訓練事例集合を生成し, それぞれから学習した弱学習器を飼育データに対する正判別率によってフィルタリングすることで, 弱学習器の採用, 不採用を決定している. このとき, 前節で説明したように, サンプリングした訓練事例の分布に偏りがある場合や, $P_N[x, c]$ に由来する訓練事例が多く含まれている場合, その訓練事例集合から学習した弱学習器は採用されにくい.

提案手法では、BaggTaming の手続きにおいて不採用となった弱学習器に着目する．この弱学習器の学習に用いられた訓練事例集合には、 $P_N[x, c]$ に由来する訓練事例と $P[x, c]$ に由来する訓練事例が混在すると考えられる．この訓練事例集合を $D_s = \{(x_i, c_i)\}_{i=1}^{N_s}$ とする．ただし、 $N_s = |D_s|$ である．ここで、特徴ベクトル x_i を $P[x, c]$ に射影し、推定ラベル \hat{c}_{rmi} を与えることができれば、目標概念に従った訓練事例 (x_i, \hat{c}_i) を得ることができる．これにより、 $P[x, c]$ に由来する訓練事例数が増加し、より多くの訓練事例を活用した学習が期待できる．

図 3.7 を用いて、提案手法の目標を説明する．図 3.7 のグラフは、正しいラベルを与えられた訓練データ数を増加させて学習した際の正判別率を表している．図 3.7 から、学習に用いた訓練事例が多くなるほど、学習精度は向上することがわかる．しかし、現実には厳密に管理された教師データである飼育データは、全データ数に比べごくわずかであり、大量の飼育データを確保することは困難である．少ない訓練事例を用いた学習では、学習精度は向上しにくいことがわかる（領域 A）．神嶌らの提案した BaggTaming は、飼育データに対する正判別率を用いて野生データをフィルタリングすることで、飼育データの判別に有効な野生データを選択的に使用している．これにより、飼育データと同水準でラベルを与えられた野生データを活用できるため、学習に用いる訓練事例数が増え、学習精度の向上を実現している（領域 B）．ここで、BaggTaming の手続きにおいて、不採用となった訓練事例は使用されない．不採用となった訓練事例に、所望する水準でラベルを与えることができれば、より多くの訓練事例を学習に用いることができると考えられる（領域 C）．提案手法の目的は、BaggTaming の手続きにおいて不採用となった訓練事例に適切なラベルを与えることで、従来手法では活用できていなかった訓練事例を活用可能にし、それを学習に取り込むことで限られたデータを最大限活用しつつ、さらなる学習精度の向上を目指すことである．



- A:Tame data
- B:Tame data + Wild data
- C:Tame data + Wild data + Relabeling wild data

図 3.7: Purpose of the proposed method

3.2.2 提案手法のアルゴリズム

提案手法の具体的なアルゴリズムを示す．まず，従来手法である BaggingTaming と同様に，野生データからブートストラップサンプリングを行い訓練事例集合 $D = \{D_1 \dots D_i\}$ を生成する． $D_1 \dots D_i$ から学習した弱学習器 $\hat{f}_1 \dots \hat{f}_i$ を作成し，それぞれの飼育データに対する正判別率 $p_1 \dots p_i$ を算出する．次に，飼育データから学習した基準学習器 f_T の飼育データに対する正判別率 p_T を計算する．本研究では，弱学習器の採用基準として p_T を使用する． $f_1 \dots f_i$ の飼育データに対する正判別率 $p_1 \dots p_i$ と p_T をそれぞれ比較し， p_T を上回る正判別率を示した弱学習器を採用する．採用された弱学習器を $f_1^a \dots f_j^a$ ，不採用となった弱学習器を $f_1^r \dots f_k^r$ とする．同様に，それぞれの弱学習器の学習に使用された訓練事例集合を $D_1^a \dots D_j^a$ ，

$D_1^r \dots D_k^r$ とする．ここで， $j + k = i$ である．

ここまでは従来手法と同様の手続きである．従来手法では，採用判定において不採用となった弱学習器 $f_1^r \dots f_k^r$ は破棄され，今後の手続きに影響を与えない．一方，提案手法では，採用判定において不採用となった弱学習器の学習に用いられた訓練事例集合 $D_1^r \dots D_k^r$ を保存しておく． $D_1^r \dots D_k^r$ に含まれるすべての訓練事例に対して，採用された弱学習器 $f_1^a \dots f_j^a$ の凝集により，与えるべきラベルを推定する．任意の訓練事例 (x_l, c_l) に対する推定ラベル \hat{c}_l は，形式的には次式のように決定される．

$$\hat{c}_l = \arg \max_{c \in C} \sum_{t=1}^j p_t I[c = f_t^a(x_l)] \quad (3.1)$$

ただし， C はクラスの定義域であり， $I[\cdot]$ は指示関数である．

選択した訓練事例が目標概念と無矛盾な基準でラベル付けされている場合， $\hat{c}_l = c_l$ となることが期待される．もし互いのラベルが異なる場合，その訓練事例は， $P_N[x, c]$ に由来する訓練事例であると考えられる．その訓練事例に対して， \hat{c}_l を新たなクラスラベルとして与える． $D_1^r \dots D_k^r$ に含まれるすべての訓練事例に対してこの処理を適用することで，新たな訓練事例集合 $\hat{D} = \{\hat{D}_1^r \dots \hat{D}_k^r\}$ を得る． \hat{D} から再度学習を行い， $\hat{f}_1 \dots \hat{f}_k$ を作成する． $\hat{f}_1 \dots \hat{f}_k$ のそれぞれから飼育データに対する正判別率 $\hat{p}_1 \dots \hat{p}_k$ を算出し， p_T と比較して採用判定を行う．新たに採用された弱学習器 $\hat{f}_1 \dots \hat{f}_m (m \leq k)$ と $f_1^a \dots f_j^a$ を凝集し，最終的な判定を行う．任意の特徴ベクトル x に対する推定ラベル \hat{c} は，形式的には次式のように決定される．

$$\hat{c} = \arg \max_{c \in C} \sum_{t=1}^j p_t I[c = f_t^a(x)] \quad \sum_{u=1}^m p_u I[c = f_u^a(x)] \quad (3.2)$$

以上の手続きを図 3.8 に示す．

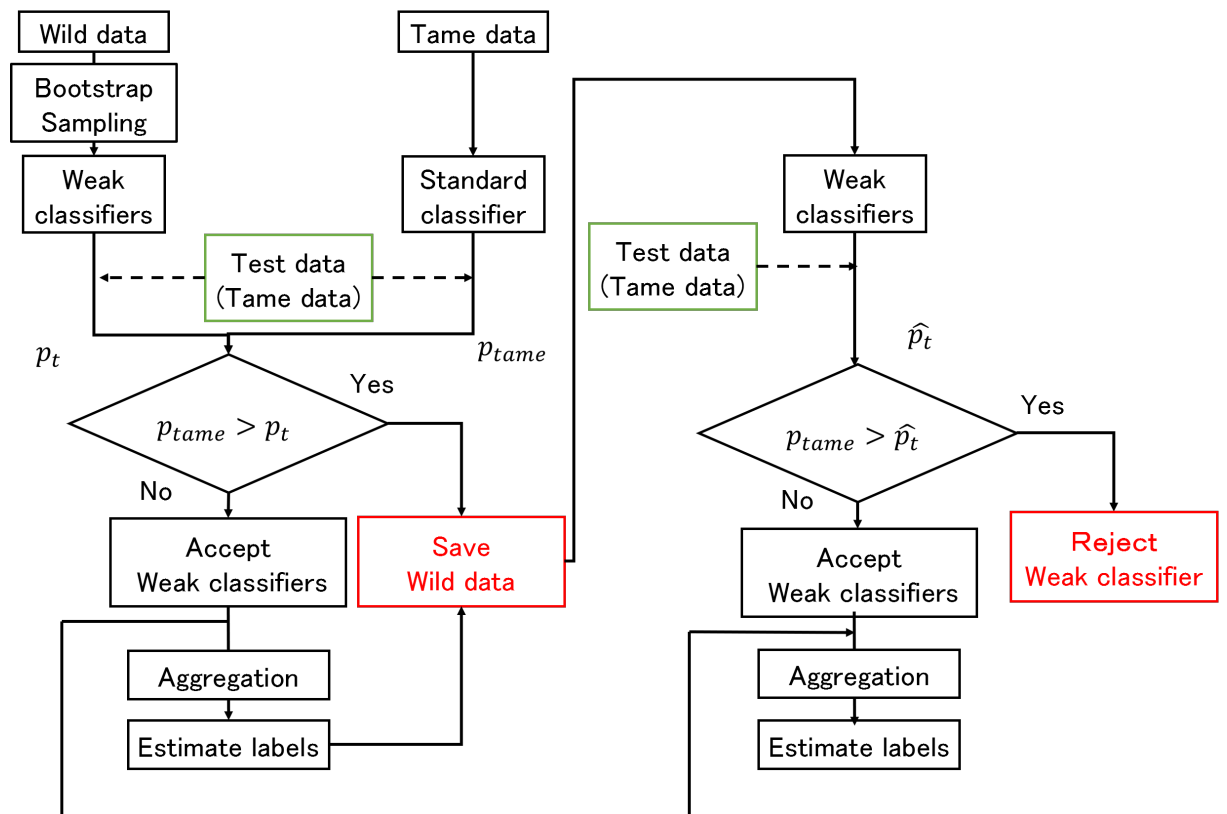


図 3.8: Algorithm of Proposed method

第4章 汎用データセットを用いた シミュレーション実験

本章では、提案手法の有効性の検証と、汎用データセットを用いたシミュレーション実験を行う。提案手法の検証では、以下の点について考察する。

1. ラベル推定精度
2. ラベル推定による野生データの推移
3. 提案手法の判別能力

それぞれの検証内容について概説する。まず、提案手法は、従来手法の手続きにおいて採用となった弱学習器を用いて、不採用となった訓練事例集合に対して与えるべきラベルを推定している。ラベル推定後の訓練事例集合から再度学習し、新たな弱学習器を加えることで、判別能力の向上を図っている。したがって、与えるべきラベルの推定精度は非常に重要であり、多くの訓練事例に対して誤ったラベルを与えてしまう場合、判別能力は悪化してしまうと考えられる。そのため、提案手法の不採用となった訓練事例に対するラベル推定精度を検証することで、提案手法の妥当性を図る。

次に、ラベル推定による野生データの推移について検証する。提案手法の手続きでは、野生データからサンプリングした弱学習器のうち、BaggTamingの手続きにおいて不採用となった弱学習器を対象としてラベルを推定している。不採用となった弱学習器には、誤ったラベルを与えられた訓練事例が多く含まれ

ていると考えられる。したがって、提案手法が有効に作用するためには、推定ラベルを与える前後で誤ったラベルを持つ訓練事例数が減少していることが重要となる。そのため、提案手法適用前後の誤ったラベルを持つ訓練事例数の推移を検討する。

最後に、提案手法の判別能力を検討する。機械学習の性能は、一般的に未知データに対する判別能力によって評価される。ここでいう未知データは、学習に用いられていない訓練事例である。従来手法である Bagging と提案手法の判別能力を測る指標として、飼育データに対する正判別率を用いて検討する。

上記の三つの項目に対し提案手法の妥当性を検討する。

4.1 シミュレーション実験に用いる汎用データセット

シミュレーション実験には、二種類のデータセットを用いる。一つは、UCI レポジトリにまとめられている汎用データセットである “Wine Quality Dataset” [92] を用いる。このデータセットは各サンプルごとに 10 種類の化学成分を計測し、ワインエキスパートによる官能検査によって品質を 10 段階に分類したラベルを持つ。もう一つは、同じく UCI レポジトリにまとめられている “Abalone Dataset” [93] を用いる。このデータセットは各サンプルごとに 9 種類の身体特徴を計測し、その年齢をまとめたデータセットである。“Wine Quality Dataset” は、官能検査に基づいてラベルが与えられている定性的なデータセットであり、“Abalone Dataset” は客観的な指標に基づいてラベルが与えられている定量的なデータセットである。これらのデータセットを用いて、シミュレーション実験を行う。

シミュレーション実験は、二値分類と多値分類に分けて検証を行う。二値分類と多値分類に対応できるよう、汎用データセットにラベルを与える。新たにラベルを与えたデータセットの概要を、それぞれ表 4.1, 表 4.2, 表 4.3, 表 4.4 に

表 4.1: Wine quality binary dataset

Original labels	Number of data	New label
5	1457	1
6	2198	2

表 4.2: Abalone binary dataset

Original labels	Number of data	New label
1~9	2096	1
10~29	2081	2

示す.

4.2 実験条件

実験では、二値分類問題では、飼育データ数 $N_T = \{40, 60, 80\}$ 、多値分類問題では $N_T = \{60, 90, 120\}$ と段階的に変化させた。また、野生データには飼育データから選ばれなかった訓練事例を用いる。したがって、飼育データと野生データは互いに素である。また、野生データには $P_N[x, c]$ に由来する訓練事例、すなわち誤ったラベルを持つ訓練事例が含まれている。その状況を再現するため、野生データに含まれる訓練事例からいくつかの訓練事例を無作為に抽出し、意図的にラベルを変更させて野生データに混入しておく。本検証では、 $P_N[x, c]$ に由来する訓練事例は、野生データの訓練事例数に依存して段階的に変化させた。 $P_N[x, c]$ に由来する訓練事例の含有率を $W_f = \{20\%, 40\%, 60\%\}$ とした。したがって、野生データに含まれる $P_N[x, c]$ に由来する訓練事例の数は、 $N_W \times W_f$

表 4.3: Wine quality multiclass dataset

Original labels	Number of data	New label
5	1457	1
6	2198	2
7	880	3

表 4.4: Abalone multiclass dataset

Original labels	Number of data	New label
1~8	1407	1
9~10	1323	2
11~29	1447	3

となる．野生データから復元抽出する訓練事例数は，使用するデータセットによって適切な値が決まるが，これは未知であるため， N_T の 50[%] とした．生成する分類器数 $i = 100$ とし，学習には決定木を用いた．基準学習器 f_T の正判別率は，飼育データを用いた LOOCV で求めた．また，判別能力は 5 分割の交差検証を用いて評価した．

4.3 不採用訓練事例に対するラベル推定精度の検証

提案手法は，BaggTaming の手続きにおいて不採用となった弱学習器に対して，その学習に用いられた訓練事例集合に対して与えられるべきラベル推定を行う．新たに獲得した野生データ \hat{D} を活用することで，学習に活用可能な訓練事例数を増加させ，判別能力の向上を目指している．ここで，提案手法によって獲得できる \hat{D} について，どれだけの訓練事例が活用可能となるかを検証する．上記の二種類のデータセットを用いて，提案手法適用前後におけるラベルの推定精度を検討する．

4.3.1 二値分類におけるラベル推定精度

まず，二値分類における提案手法適用前後のラベル推定精度を検証する．

“Wine Quality Dataset” を用いた結果を図 4.1 に示す．図 4.1 は，不採用訓練事例として抽出された訓練事例に対して，真に与えるべきラベルを正確に与えた正答率を表している．図 4.1 から，飼育データ数 N_T が大きくなるにつれて，ラ

ベルの正答率は向上していることがわかる．これは， N_T が大きくなるにつれ，採用される弱学習器の判別能力が向上するためであると考えられる．弱学習器の採用判定には，飼育データから学習している基準学習器の飼育データに対する正判別率を用いている．そのため， N_T が大きくなるにつれて，基準学習器の判別能力が高くなり，それに伴い採用される弱学習器の判別能力も向上しているため，ラベル推定の精度が向上していると考えられる．

“Abalone Dataset”を用いた結果を図 4.2 に示す．図 4.2 からわかるように，“Abalone Dataset”を用いた結果も“Wine Quality Dataset”を用いた場合と同様に，飼育データ数 N_T が大きくなるにつれて，ラベルの正答率は向上していることがわかる．また， W_f が大きくなるにつれ，微差ではあるがラベル推定の正答率は下がる傾向にあることがうかがえる．これは， W_f が大きくなるにつれて，野生データ中に含まれる $P_N[x, c]$ に由来する訓練事例数が増えることに起因すると考えられる． W_f の変動による効果は二つ考えられる． W_f が大きくなるほど，ブートストラップサンプリングを用いて作成した訓練事例集合には $P_N[x, c]$ に由来する訓練事例が含まれやすくなると考えられる．そのため，それらから学習した弱学習器は飼育データに対する判別能力にばらつきが生じるため，採用されにくくなると考えられる．その一方で，アンサンブル学習では多様な弱学習器の凝集によって判別精度が向上することも報告されている [84]．したがって，弱学習器の採用数の減少と採用弱学習器の多様性の向上の二つの効果の結果，緩やかな減少傾向を生み出したと考えられる．

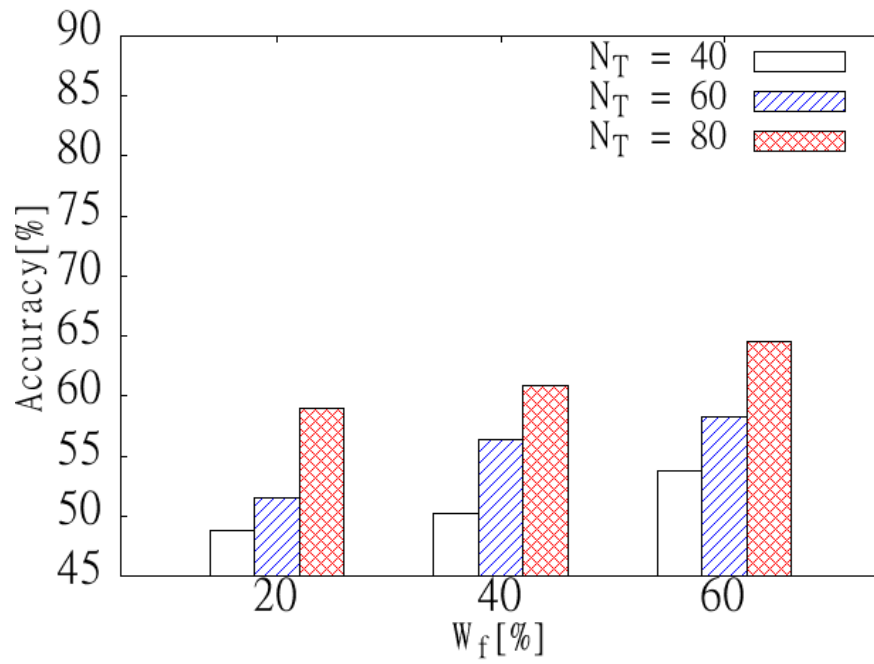


図 4.1: Estimation accuracy of Binary label using “Wine Quality Dataset”

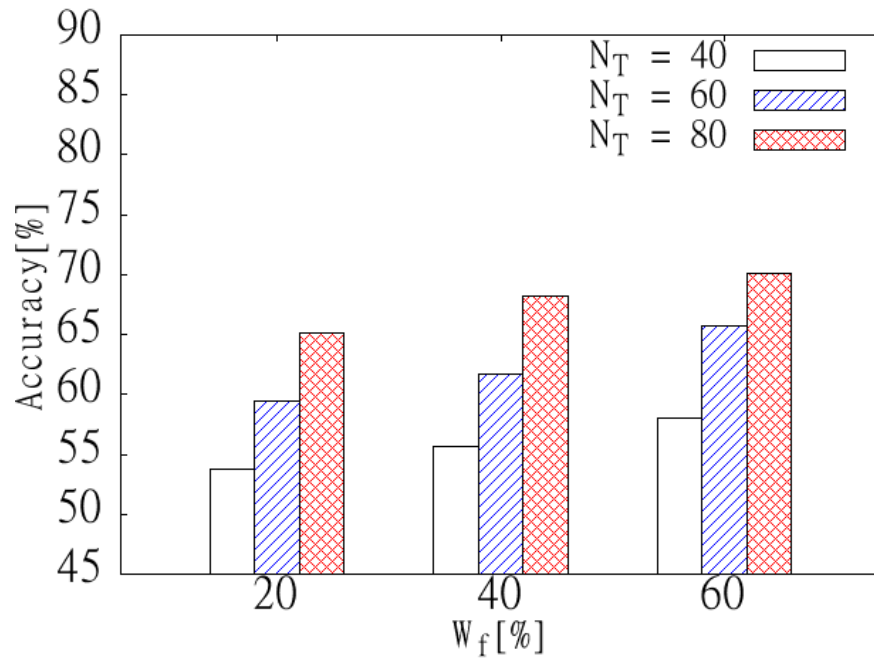


図 4.2: Estimation accuracy of Binary label using “Abalone Dataset”

4.3.2 多値分類におけるラベル推定精度

多値分類における提案手法適用前後のラベル推定精度を検証する．本研究では，三値分類問題を対象とした検証を行う．

“Wine Quality Dataset”を用いた結果を図4.3に示す．図4.3は，不採用訓練事例として抽出された訓練事例に対して，真に与えるべきラベルを正確に与えた正答率を表している．図4.3から，二値分類と同様に，飼育データ数 N_T が大きくなるにつれて，ラベルの正答率は向上していることがわかる．これは， N_T が大きくなるにつれ，採用される弱学習器の判別能力が向上するためであり，多値分類においても二値分類と同様であると考えられる．弱学習器の採用判定には，飼育データから学習している基準学習器の飼育データに対する正判別率を用いている．そのため， N_T が大きくなるにつれて，基準学習器の判別能力が高くなり，それに伴い採用される弱学習器の判別能力も向上しているため，ラベル推定の精度が向上していると考えられる．

“Abalone Dataset”を用いた結果を図4.4に示す．図4.4からわかるように，“Abalone Dataset”を用いた結果も“Wine Quality Dataset”を用いた場合と同様に，飼育データ数 N_T が大きくなるにつれて，ラベルの正答率は向上していることがわかる．また，二値分類問題と比較して，多値分類問題は N_T が大きくなるにつれ，ラベルの正答率は大きく変動している．これは，二値から三値へと分類対象が増えたことに起因していると考えられる．一般的な機械学習は，分類対象が増えるにつれ判別は困難となるため，より多くの訓練事例から学習する必要が生じる．そのため， N_T が小さい場合，飼育データに対する判別能力は小さくなり， N_T の増加に伴い，判別能力が大きく向上したと考えられる．

また，多値分類問題では， W_f が大きくなるにつれ，ラベル推定の正答率は下がる傾向にあることがわかる．これは，二値分類と同様に， W_f が大きくなるにつれて，野生データ中に含まれる $P_N[x, c]$ に由来する訓練事例数が増えることに

起因すると考えられる。 W_f が大きくなるほど、ブートストラップサンプリングを用いて作成した訓練事例集合には $P_N[x, c]$ に由来する訓練事例が含まれやすくなると考えられる。それにより、弱学習器の採用数が減少したと考えられる。また、多値分類では、二値分類と比較して判別境界が複雑になりやすい。そのため、所望する分類基準を満たす判別境界のマージンは二値分類よりも限定されてしまう。本研究では飼育データに対する判別能力を用いて弱学習器の採用判定を行っているため、高い判別能力を示す弱学習器は二値分類と比較してばらつきが低減していると考えられる。つまり、多様な弱学習器を採用することは困難であり、それによりラベル推定に用いる学習器の判別能力が向上できなかったと考えられる。

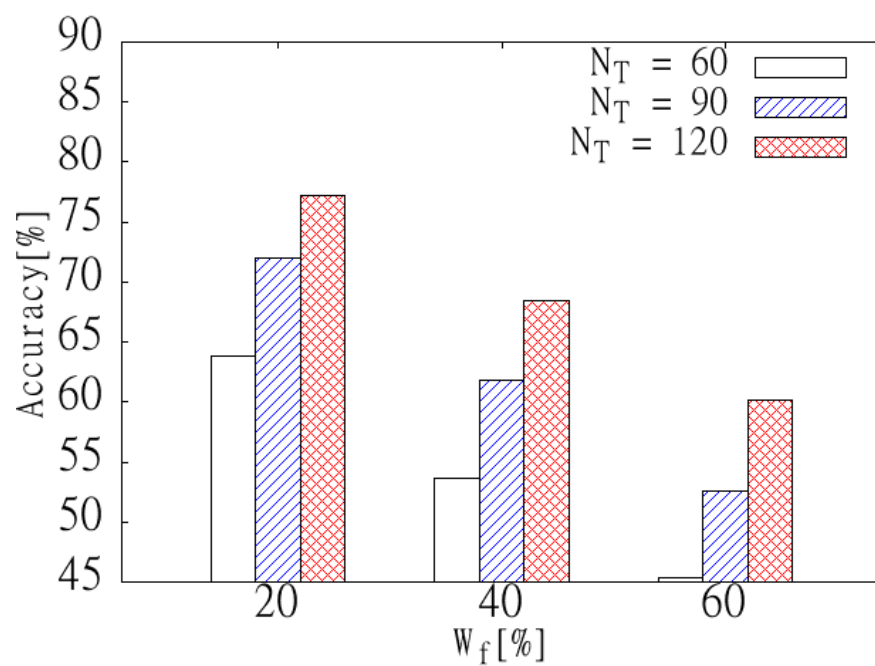


図 4.3: Estimation accuracy of multilevel label using “Wine Quality Dataset”

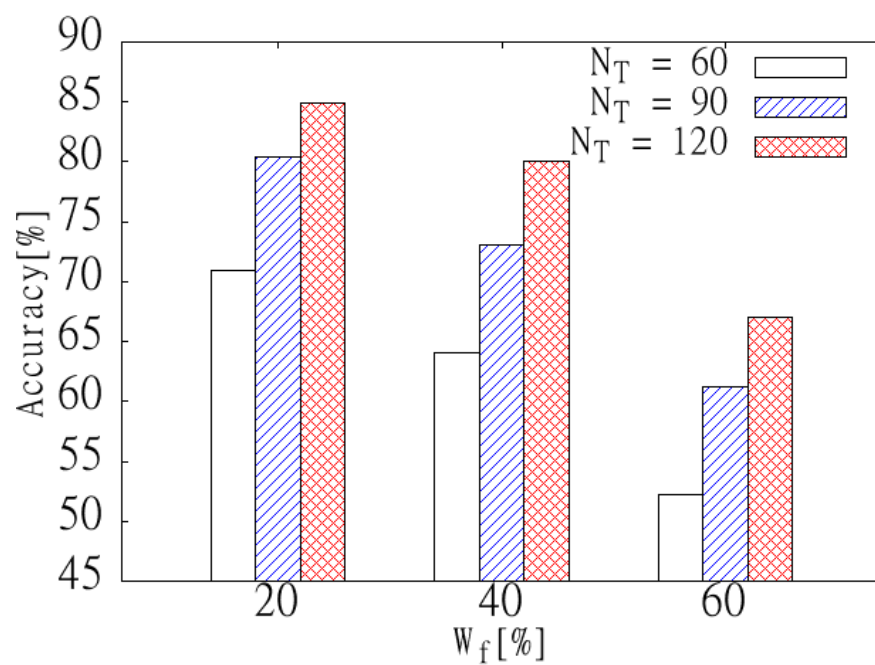


図 4.4: Estimation accuracy of multilevel label using “Abalone Dataset”

4.4 ラベル推定による野生データの推移

二つ目の検証として、提案手法の適用前後の野生データの推移を検討する．提案手法がうまく作用すれば、野生データ中にある $P_N[x, c]$ に由来する訓練事例を削減することができ、より多くの訓練事例を用いた学習が可能となる．具体的な検証方法は、提案手法適用前後の野生データを調査し、与えられるべきラベルを持つ訓練事例の減少数および減少率を検証する．

4.4.1 二値分類における野生データの推移

まず、二値分類における提案手法適用前後の野生データの推移を検証する．

“Wine Quality Dataset”を用いた場合の野生データの減少率を図4.5に示す．図4.5は提案手法適用前の野生データに含まれている誤ったラベルを持つ訓練事例と、提案手法適用後野生データに含まれている誤ったラベルを持つ訓練事例の減少率である．また、提案手法適用前後の野生データに含まれる誤ったラベルを持つ訓練事例の数を表4.5に示す．

図4.5から、すべての条件で誤ったラベルを持つ訓練事例数が減少していることがわかる．また、その減少率は飼育データ数 N_T が増加するにつれて向上していることがわかる．これは、4.3で示したラベル推定の精度に依存していると考えられる．ラベル推定の精度が高くなるほど、与えられるべきラベルを正しく判定できる．また、本検証では、不採用となった訓練事例 \hat{D} に対してラベル推定を行っている． \hat{D} には $P_N[x, c]$ に由来する訓練事例と、 $P[x, c]$ に由来する訓練事例が混在しているため、 $P[x, c]$ に由来する訓練事例に対してもラベルを推定している．したがって、 $P_N[x, c]$ に由来する訓練事例に対して真に与えられるべきラベルを推定できた場合のみ、誤った訓練事例の数は減少する．そのため、ラベル推定精度よりも低い減少率となると考えられる．

また、表 4.5 から、 W_f が大きくなるほど、誤ったラベルを持つ訓練事例を適切に修正しやすくなる。これは、 W_f が大きいほどサンプリングされた訓練事例集合の中に誤ったラベルを持つ訓練事例が含まれやすくなることが主な要因であると考えられる。

“Abalone Dataset” を用いた場合の野生データの減少率を図 4.6 に示す。また、提案手法適用前後の野生データに含まれる誤ったラベルを持つ訓練事例の数を表 4.6 に示す。

図 4.6 からわかるように、“Abalone Dataset” も “Wine Quality Dataset” と同様に、すべての条件で誤ったラベルを持つ訓練事例が減少していることがわかる。また、その減少率は飼育データ数 N_T が増加するにつれて向上していることがわかる。

“Wine Quality Dataset” の減少率 (図 4.5) と “Abalone Dataset” の減少率 (図 4.6) を比較する。 $N_T = 40$ の場合の減少率は、ラベル推定精度は “Abalone Dataset” のほうが高いにもかかわらず、“Abalone Dataset” のほうが減少率は低くなっている。これは、“Abalone Dataset” を用いた実験で不採用となった訓練事例集合に含まれている誤ったラベルを持つ訓練事例数が少ないことが原因であると考えられる。

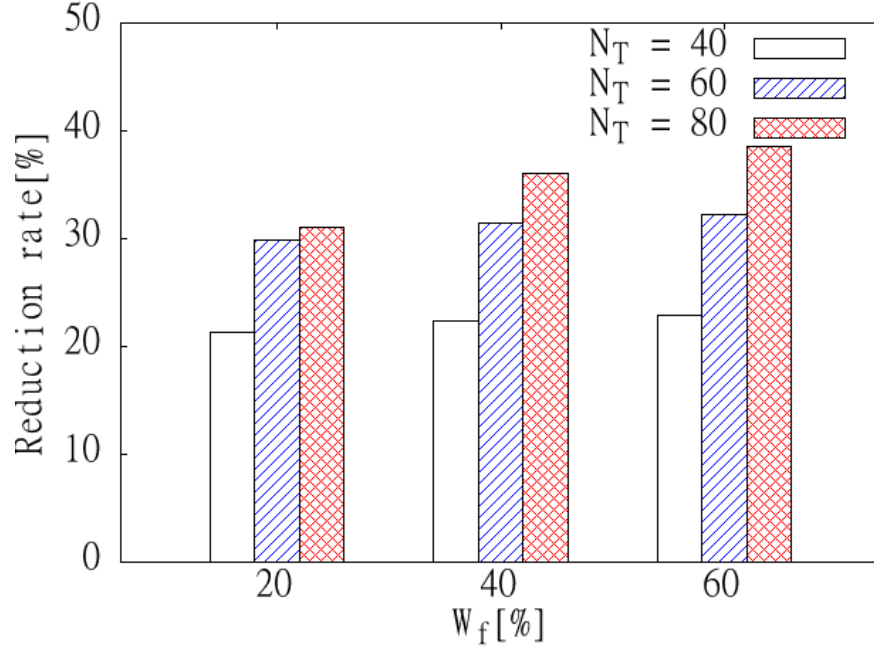


図 4.5: Reduction rate of training examples in binary classification using “Wine Quality Dataset”

表 4.5: Reduction rate of binary classification using “Wine Quality dataset”,

		N_T					
		40		60		80	
		Before	After	Before	After	Before	After
W_f	20	722	567.6	718	503.7	714	491.9
	40	1446	1123.4	1438	985.6	1430	914.0
	60	2168	1670.8	2156	1460.4	2144	1316.5

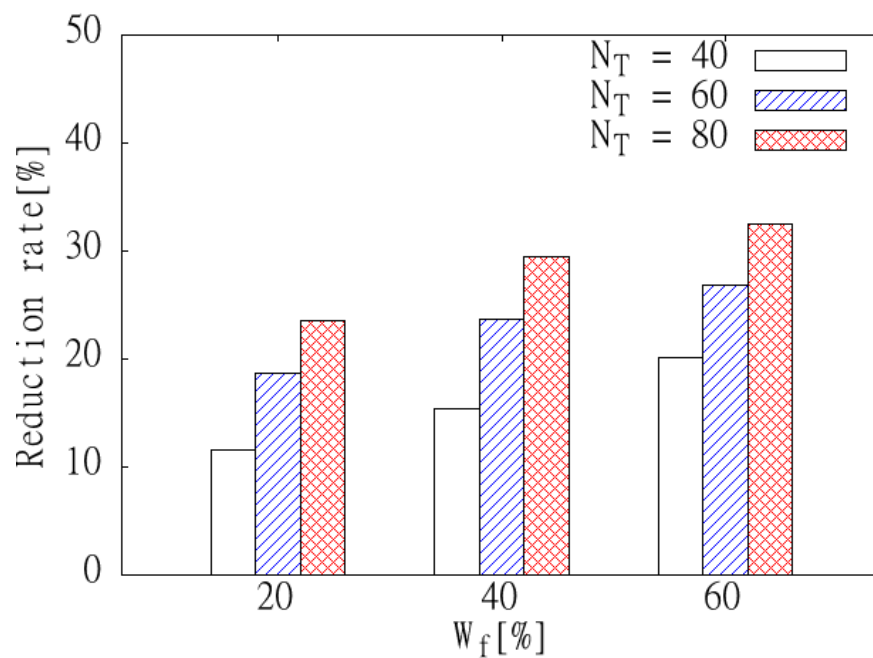


図 4.6: Reduction rate of training examples in binary classification using “Abalone Dataset”

表 4.6: Reduction rate of binary classification using “Abalone dataset”

		N_T					
		40		60		80	
		Before	After	Before	After	Before	After
W_f	20	826	730.6	822	668.9	818	625.7
	40	1654	1399.1	1646	1255.1	1638	1155.9
	60	2482	1981.9	2470	1808.4	2458	1660.6

4.4.2 多値分類における野生データの推移

次に、多値分類における提案手法適用前後の野生データの推移を検証する。

“Wine Quality Dataset”を用いた場合の野生データの減少率を図4.7に示す。図4.7は提案手法適用前の野生データに含まれている誤ったラベルを持つ訓練事例と、提案手法適用後の野生データに含まれている誤ったラベルを持つ訓練事例の減少率である。また、提案手法適用前後の野生データに含まれる誤ったラベルを持つ訓練事例の数を表4.7に示す。

“Abalone Dataset”を用いた場合の野生データの減少率を図4.8に示す。また、提案手法適用前後の野生データに含まれる誤ったラベルを持つ訓練事例の数を表4.8に示す。

図4.7と図4.8からわかるように、“Abalone Dataset”と“Wine Quality Dataset”はともに、すべての条件で誤ったラベルを持つ訓練事例数が減少していることがわかる。また、その減少率は二値分類よりも高いことがわかる。これは、二値分類と比べ、多値分類では不採用となる弱学習器が多く、結果的に多くの誤った訓練事例に対してラベルを推定したことが下人であると考えられる。

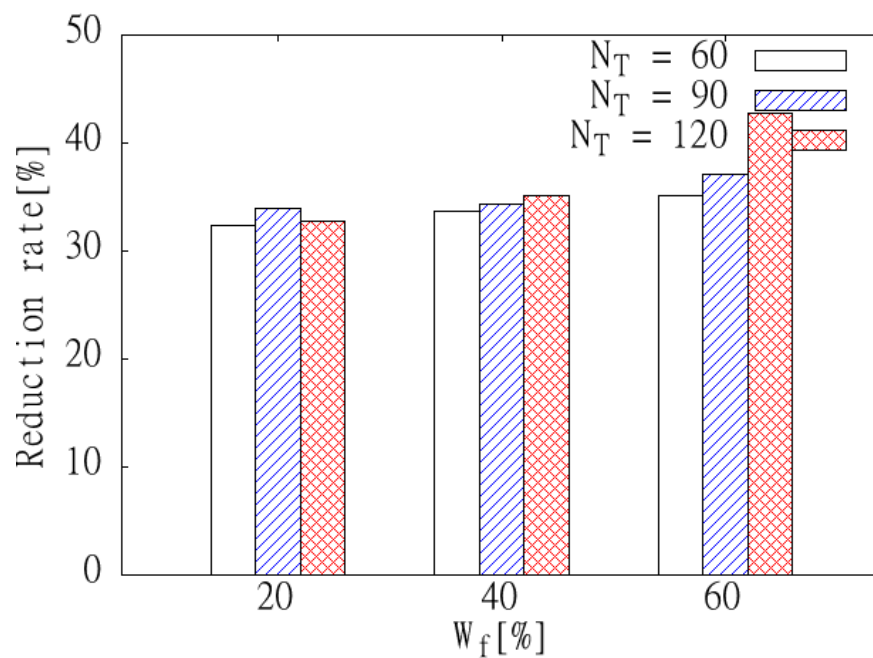


図 4.7: Reduction rate of training examples in multilevel classification using “Abalone Dataset”

表 4.7: Reduction number of multilevel classification using “Wine Quality dataset”

		N_T					
		60		90		120	
		Before	After	Before	After	Before	After
W_f	20	894	604.8	888	586.3	882	593.3
	40	1788	1186.6	1776	1166.6	1764	1145.1
	60	2220	1270.5	2203	1384.4	2186	1417.2

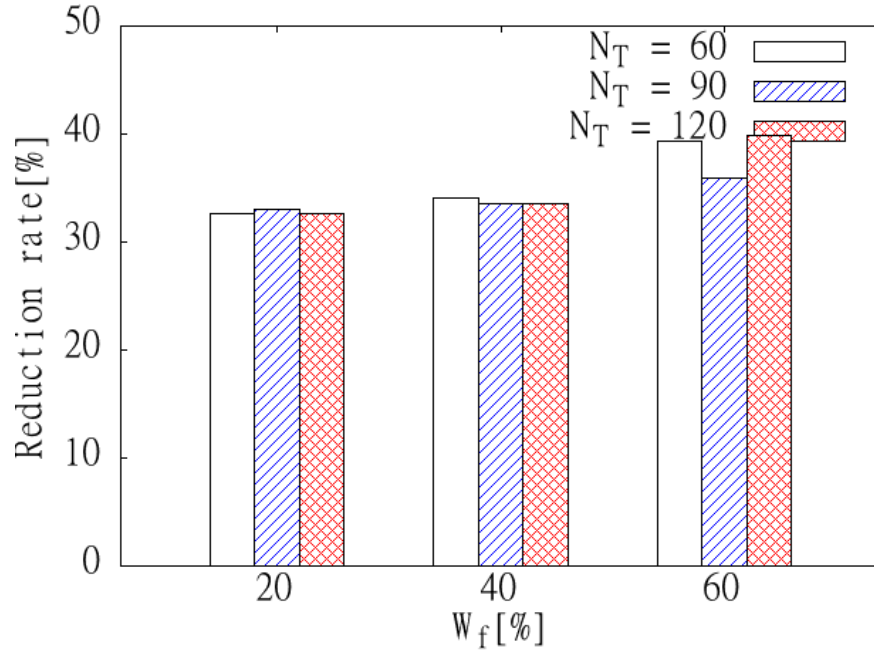


図 4.8: Reduction rate of training examples in multilevel classification using “Abalone Dataset”

表 4.8: Reduction number of multilevel classification using “Abalone dataset”

		N_T					
		60		90		120	
		Before	After	Before	After	Before	After
W_f	20	822	553.9	816	546.9	810	546.2
	40	1644	1084.8	1632	1085.0	1620	1076.0
	60	2469	1497.0	2451	1572.0	2433	1463.4

4.5 提案手法の判別能力

三つ目の検証として、提案手法の判別能力を検証する。提案手法の判別能力の比較対象として、飼育データのみから学習した Bagging と、従来手法の BaggingTaming の結果を併せて示す。

4.5.1 二値分類における判別能力

二値分類における提案手法適用前後のラベル推定精度を検証する。“Wine Quality Dataset”を用いた結果を図 4.9 から図 4.11 に示す。図 4.9, 図 4.10, 図 4.11 はそれぞれ $N_T = 40, 60, 80$ の場合を表す。また、その正判別率を表 4.9, 表 4.10, 表 4.11 に示す。

図 4.9, 図 4.10, 図 4.11 のそれぞれにおいて、Bagging は飼育データに対して 60[%] から 70[%] の判別能力を示している。これは、二値分類問題であることを考慮すると、高い精度とは言えない。このことから、少ない訓練事例からの学習では高い判別能力を示すことはできないということがわかる。

次に、従来手法の判別能力を考察する。飼育データと野生データを併用する従来手法は、飼育データに対して 70[%] から 80[%] 程の判別能力を示しており、飼育データのみを用いて学習している Bagging よりも高い判別能力を示している。この判別能力の向上から、野生データから選択的に訓練事例を用いることにより、学習に用いる訓練事例数を増加させることが有効であることがわかる。

さらに、提案手法は従来手法よりも高い判別能力を示している。4.4 で示したように、提案手法を用いることで、誤ったラベルを持つ訓練事例の数を低減させることができる。つまり、従来手法よりもさらに多くの訓練事例を学習に活用することができたため、判別能力は向上したと考えられる。また、本検証では、提案手法で使用した訓練事例集合は従来手法で生成した訓練事例集合と同

一の組合せを用いており、新たにサンプリングをしていない。したがって、判別能力の増加は学習に活用可能な弱学習器数が増えたことに起因し、したがって野生データのラベルを推定することで活用可能な訓練事例数を増加させたということになる。また、今まで採用されなかった訓練事例を学習に用いることで、アンサンブル学習の多様性の向上にも寄与したと考えられる。このことから、提案手法を用いることで、学習に活用可能な訓練事例を生成することが可能であるといえる。

“Abalone Dataset”を用いた結果を図4.12から図4.14に示す。図4.12, 図4.13, 図4.14はそれぞれ $N_T = 40, 60, 80$ の場合を表す。また、その正判別率を表4.12, 表4.13, 表4.14に示す。

図4.12, 図4.13, 図4.14のそれぞれにおいて、Baggingは飼育データに対して60[%]から70[%]の判別能力を示している。“Abalone Dataset”を用いた場合も、“Wine Quality Dataset”と同様に、Baggingは飼育データに対して60[%]から70[%]の判別能力を示している。また、飼育データと野生データを併用する従来手法は、飼育データに対して70[%]から80[%]の判別能力を示しており、飼育データのみを用いて学習しているBaggingよりも高い判別能力を示している。さらに、提案手法は従来手法よりも高い判別能力を示している。

以上のことから、二値分類において、提案手法は有効であることがわかる。

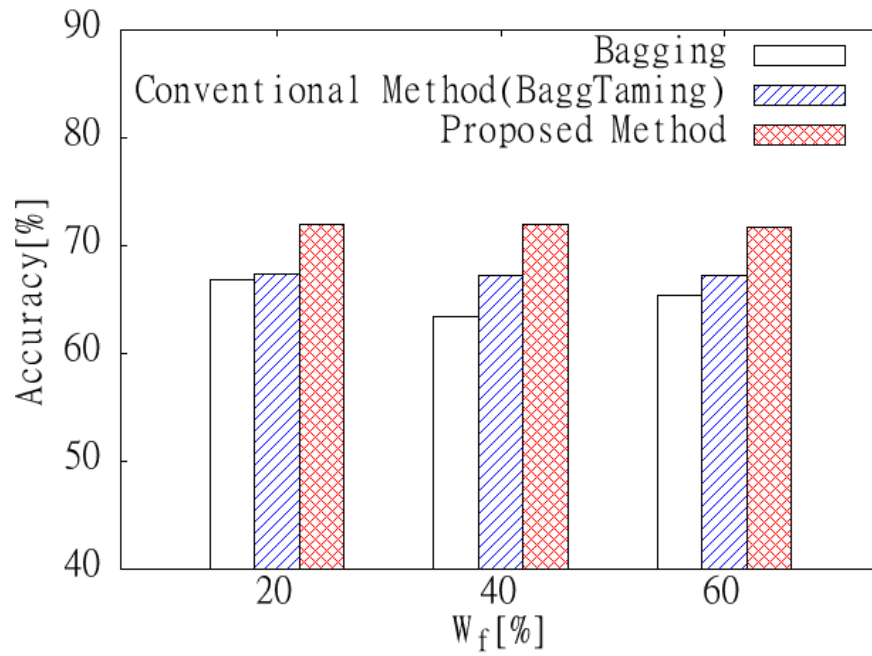


図 4.9: Accuracy of binary classification for test data using “Wine Quality Dataset”, $N_T = 40$

表 4.9: Accuracy of binary classification for test data using “Wine Quality Dataset”, $N_T = 40$

		Accuracy		
		Bagging	BaggTaming	Proposed method
W_f	20	50.15	57.99	62.78
	40	50.04	55.87	60.87
	60	50.78	52.40	58.73

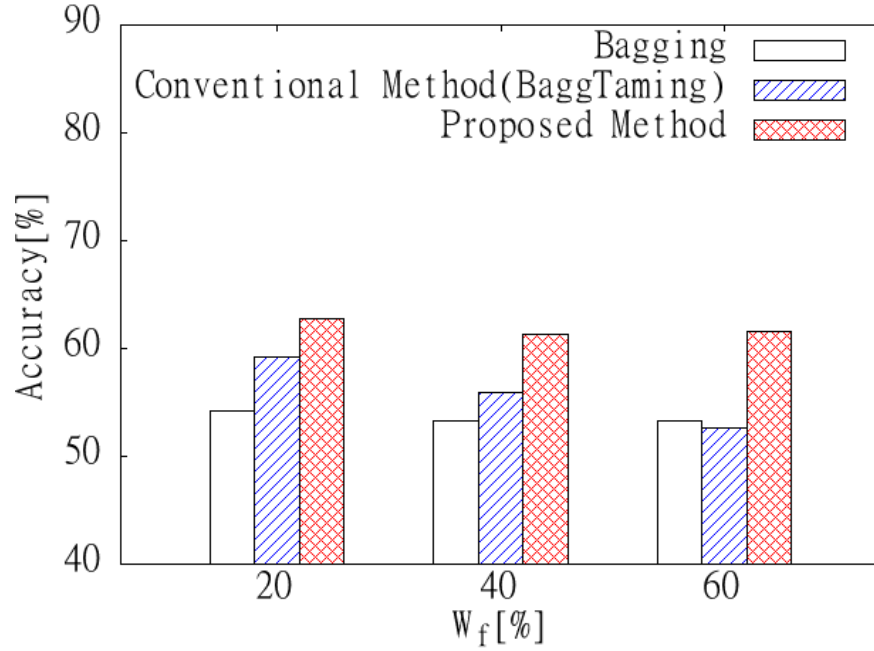


図 4.10: Accuracy of binary classification for test data using “Wine Quality Dataset”, $N_T = 60$

表 4.10: Accuracy of binary classification for test data using “Wine Quality Dataset”, $N_T = 60$

		Accuracy		
		Bagging	BaggTaming	Proposed method
W_f	20	54.24	59.21	62.71
	40	53.24	55.96	61.31
	60	53.23	52.58	61.56

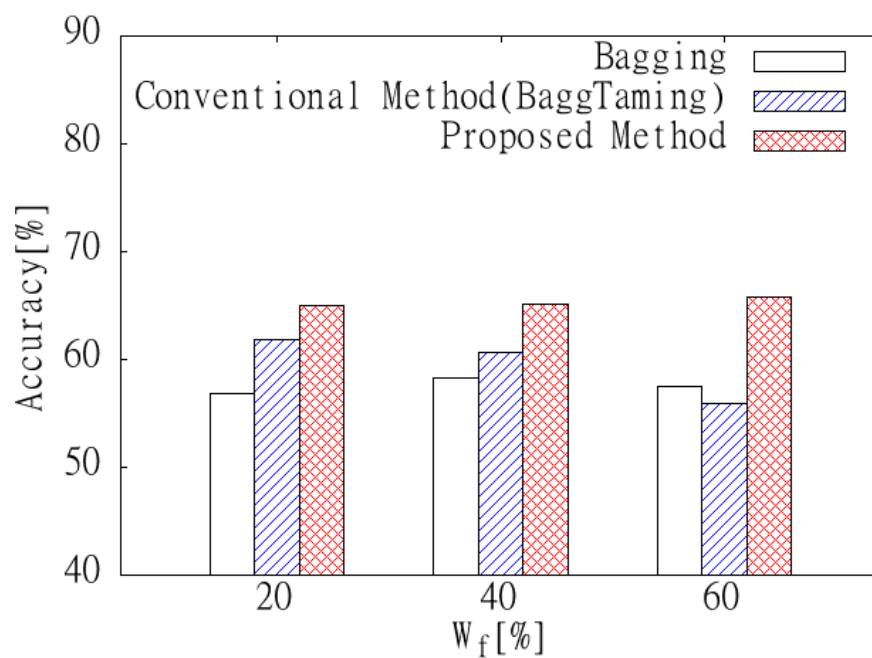


図 4.11: Accuracy of binary classification for test data using “Wine Quality Dataset”, $N_T = 80$

表 4.11: Accuracy of binary classification for test data using “Wine Quality Dataset”, $N_T = 80$

		Accuracy		
		Bagging	BaggTaming	Proposed method
W_f	20	56.80	61.83	64.96
	40	58.30	60.68	65.10
	60	57.49	55.97	65.80

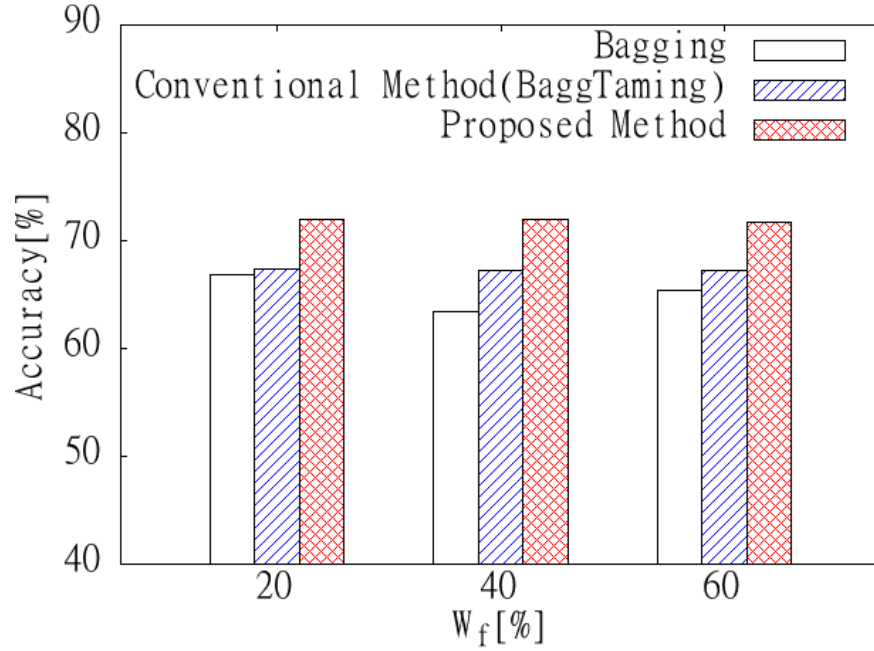


図 4.12: Accuracy of binary classification for tame data using “Abalone Dataset”, $N_T = 40$

表 4.12: Accuracy of binary classification for test data using “Abalone Dataset”, $N_T = 40$

		Accuracy		
		Bagging	BaggTaming	Proposed method
W_f	20	66.87	67.43	72.03
	40	63.36	67.18	71.92
	60	65.39	67.28	71.73

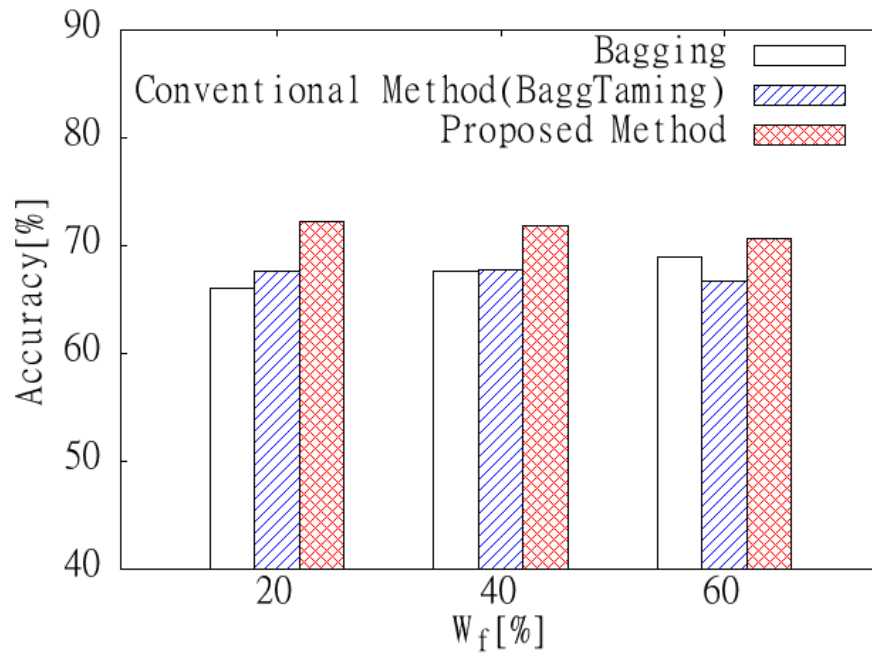


図 4.13: Accuracy of binary classification for tame data using “Abalone Dataset”, $N_T = 60$

表 4.13: Accuracy of binary classification for test data using “Abalone Dataset”, $N_T = 60$

		Accuracy		
		Bagging	BaggTaming	Proposed method
W_f	20	66.02	67.69	72.24
	40	67.60	67.78	71.82
	60	69.01	66.77	70.70

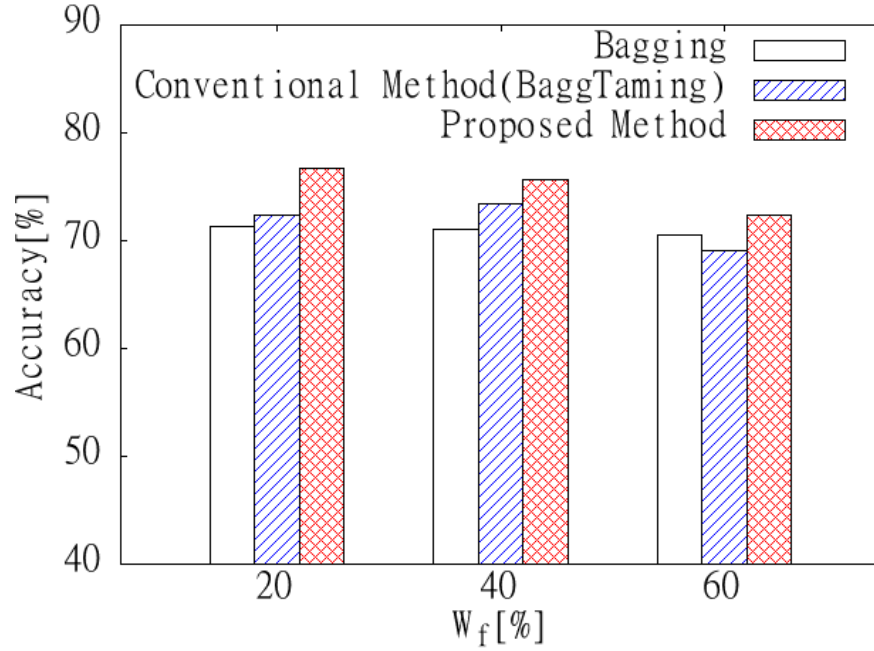


図 4.14: Accuracy of binary classification for test data using “Abalone Dataset”, $N_T = 80$

表 4.14: Accuracy of binary classification for test data using “Abalone Dataset”, $N_T = 80$

		Accuracy		
		Bagging	BaggTaming	Proposed method
W_f	20	71.38	72.36	76.69
	40	71.09	73.37	75.63
	60	70.56	69.02	72.34

4.5.2 多値分類におけるラベル推定精度

多値分類における提案手法適用前後のラベル推定精度を検証する．“Wine Quality Dataset”を用いた結果を図 4.15 から図 4.17 に示す．図 4.15, 図 4.16, 図 4.17 はそれぞれ $N_T = 60, 90, 120$ の場合を表す．その正判別率を表 4.15, 表 4.16, 表 4.17 に示す．

また, “Abalone Dataset”を用いた結果を図 4.18 から図 4.20 に示す．図 4.18, 図 4.19, 図 4.20 はそれぞれ $N_T = 60, 90, 120$ の場合を表す．その正判別率を表 4.18, 表 4.19, 表 4.20 に示す．

多値分類では, 飼育データのみを用いた Bagging は 50[%] を下回る正判別率を示すこともあり, 十分な判別能力を示すことができないということがわかる．また, 従来手法の BaggTaming では, Bagging よりも高い判別能力を示している．この判別能力の向上から, 野生データから選択的に訓練事例を用いて, 学習に用いる訓練事例数を増加させることが有効であることがわかる．さらに, 提案手法は従来手法よりも高い判別能力を示している．

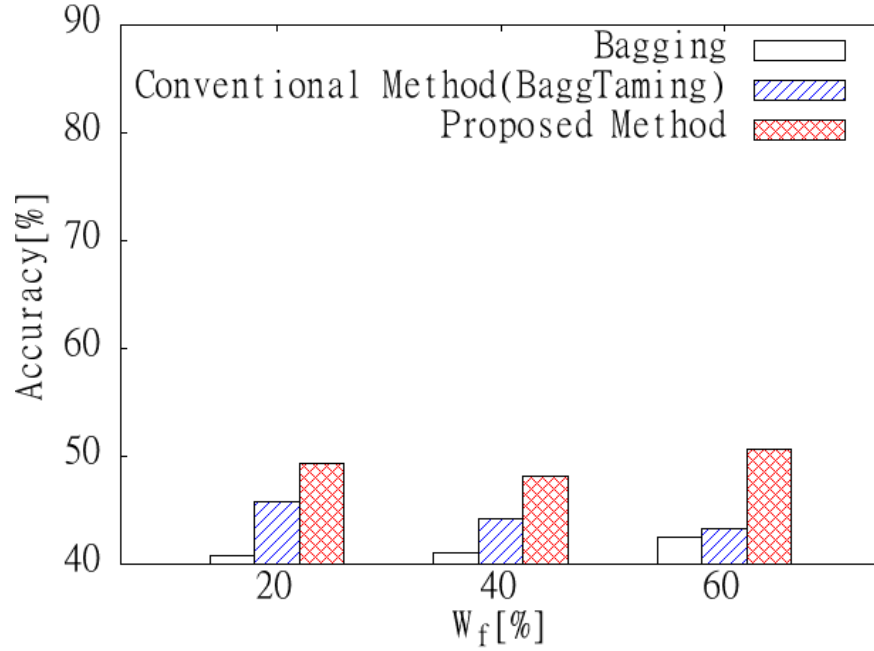


図 4.15: Accuracy of multilevel classification for test data using “Wine Quality Dataset”, $N_T = 60$

表 4.15: Accuracy of multilevel classification for test data using “Wine Quality Dataset”, $N_T = 60$

		Accuracy		
		Bagging	BaggTaming	Proposed method
W_f	20	40.73	45.85	49.32
	40	40.99	44.16	48.11
	60	42.54	43.23	50.69

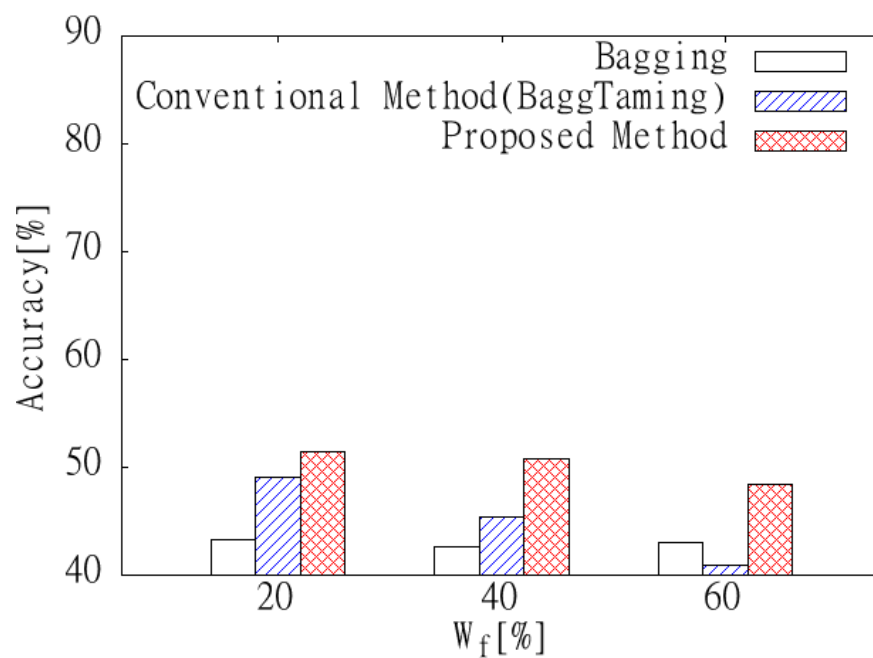


図 4.16: Accuracy of multilevel classification for test data using “Wine Quality Dataset”, $N_T = 90$

表 4.16: Accuracy of multilevel classification for test data using “Wine Quality Dataset”, $N_T = 90$

		Accuracy		
		Bagging	BaggTaming	Proposed method
W_f	20	43.30	49.07	51.43
	40	42.60	45.44	50.80
	60	43.04	40.88	48.47

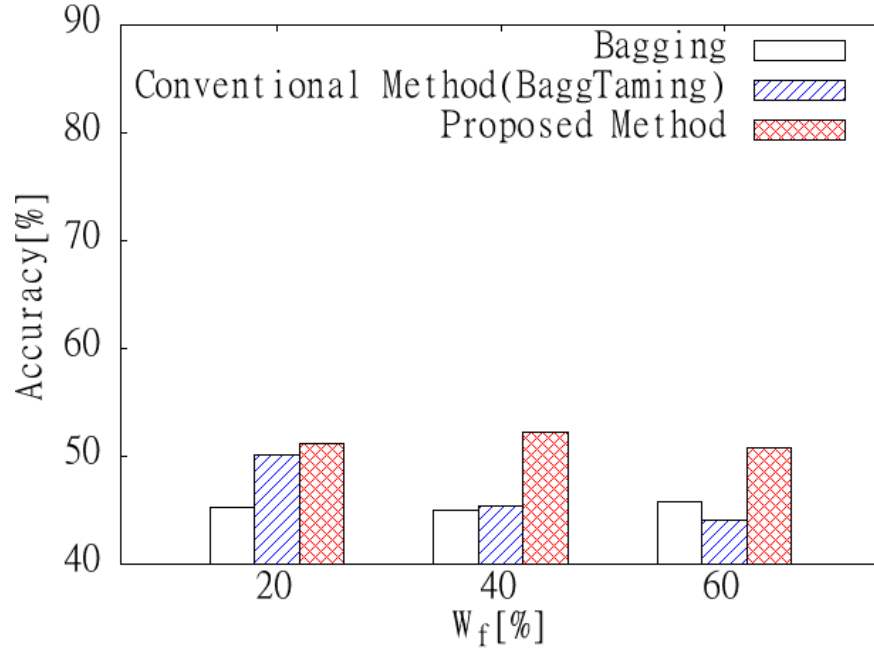


図 4.17: Accuracy of multilevel classification for test data using “Wine Quality Dataset”, $N_T = 120$

表 4.17: Accuracy of multilevel classification for test data using “Wine Quality Dataset”, $N_T = 120$

		Accuracy		
		Bagging	BaggTaming	Proposed method
W_f	20	45.20	50.13	51.22
	40	45.06	45.43	52.19
	60	45.82	44.04	50.77

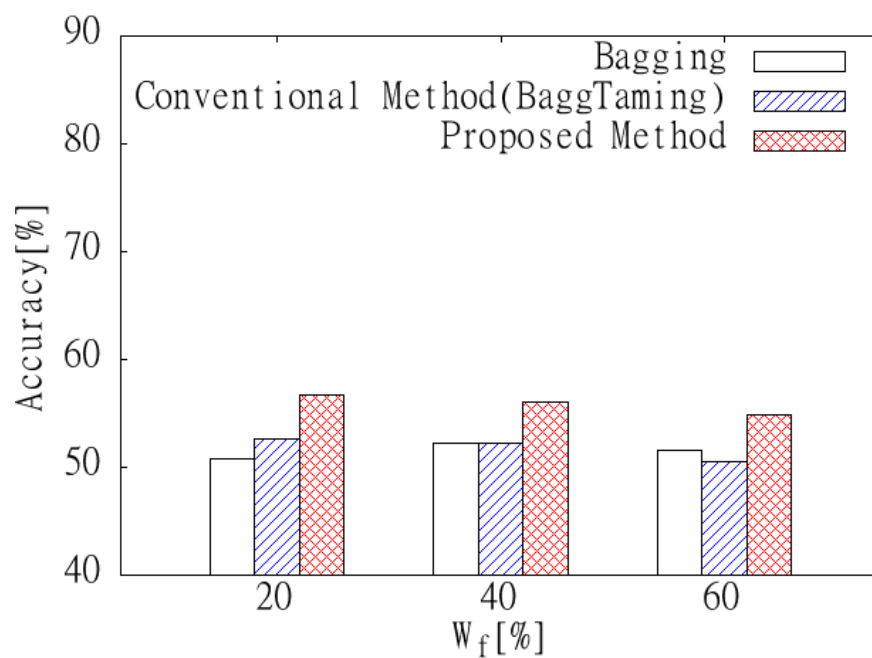


図 4.18: Accuracy of multilevel classification for test data using “Abalone Dataset”, $N_T = 60$

表 4.18: Accuracy of multilevel classification for test data using “Abalone Dataset”, $N_T = 60$

		Accuracy		
		Bagging	BaggTaming	Proposed method
W_f	20	50.82	52.60	56.68
	40	52.18	52.24	56.00
	60	51.52	50.56	54.92

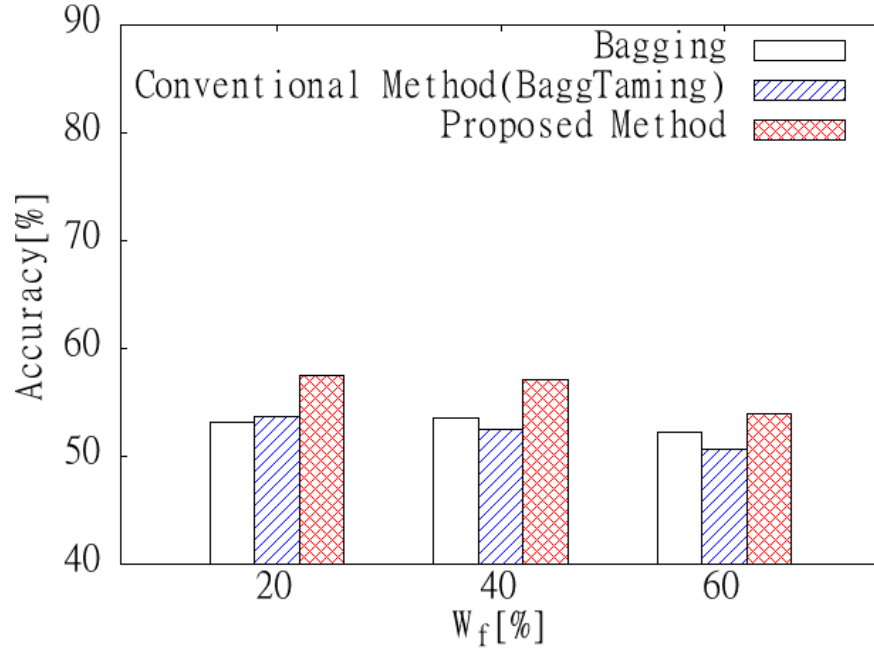


図 4.19: Accuracy of multilevel classification for test data using “Abalone Dataset”, $N_T = 90$

表 4.19: Accuracy of multilevel classification for test data using “Abalone Dataset”, $N_T = 90$

		Accuracy		
		Bagging	BaggTaming	Proposed method
W_f	20	53.10	53.74	57.56
	40	53.49	52.50	57.09
	60	52.26	50.65	53.95

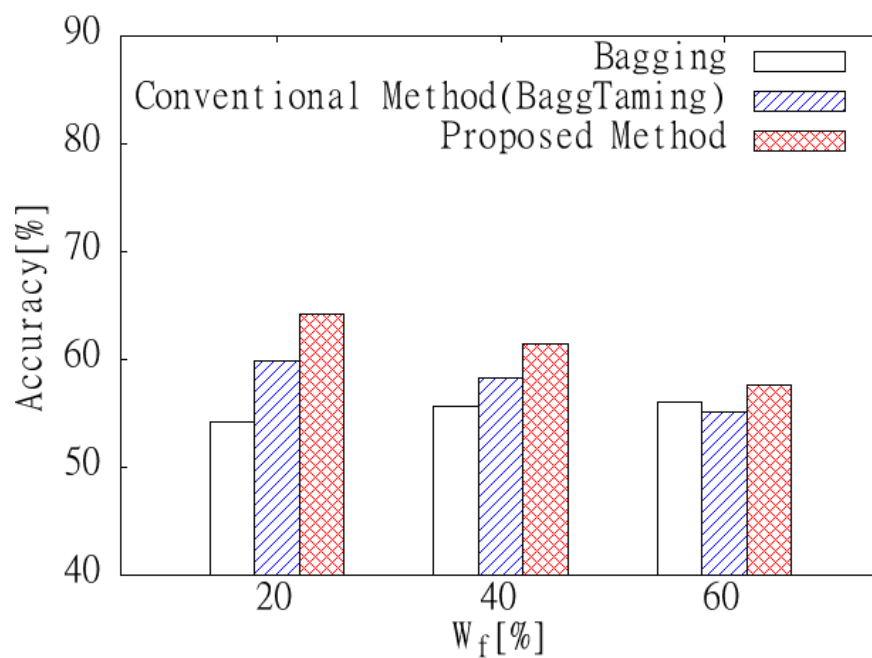


図 4.20: Accuracy of multilevel classification for test data using “Abalone Dataset”, $N_T = 120$

表 4.20: Accuracy of multilevel classification for test data using “Abalone Dataset”, $N_T = 120$

		Accuracy		
		Bagging	BaggTaming	Proposed method
W_f	20	54.27	59.93	64.15
	40	55.64	58.24	61.50
	60	56.05	55.12	57.69

4.5.3 提案手法の考察

4.5.1, 4.5.2 の結果から, 提案手法は従来手法よりも高い正判別率を示すことが分かった. ここで, 提案手法の目標について考察する. しかし, 提案手法の正判別率は二値分類では約 70[%], 多値分類では約 50[%] と低い値を示した. この原因について考察する.

本実験では, 汎用データセットに対して二値分類, 多値分類に対応するように新たにラベルを与えている. 二値分類では, “Wine quality dataset” は表 4.1, “Abalone dataset” は表 4.2 に示すようにラベルを与えている. また, 多値分類では “Wine quality dataset” は表 4.3, “Abalone dataset” は表 4.4 に示すようにラベルを与えている. 表 4.1, 表 4.3 からわかるように, “Wine quality dataset” で使用しているデータは original label が近いことがわかる. “Wine quality dataset” における original label は, ワインエキスパートの官能検査による 10 段階の味品質であるため, original label が近いデータ同士では相互に検査バイアスがかかる恐れがある. そのため, 分類が困難であり, 正判別率が低くなったと考えられる. また, “Abalone dataset” では, 表 4.2, 表 4.4 に示すような分類をしている. “Abalone dataset” の度数分布について, 二値分類を図 4.21, 多値分類を図 4.22 に示す. 図 4.21, 図 4.22 からわかるように, 本研究の分類方法では, 各クラスのうち最も多い original label は分類境界に近いデータとなっている. さらに, “Abalone dataset” はアワビの身体的特徴とその年輪によって分類されており, 個体差などによるばらつきを含むと考えられる. そのため, original label の近傍では同じようなデータが含まれていると推測できる. これらの理由により, 分類境界付近のデータが多く, さらにそれらは近傍のデータと相関が高いことが推測できる. したがって, 分類境界近傍の判定ミスが主な原因となり, 正判別率が低下したと考えられる.

表 4.21 に, 従来手法と提案手法の平均正判別率と, すべてのラベルが正しい

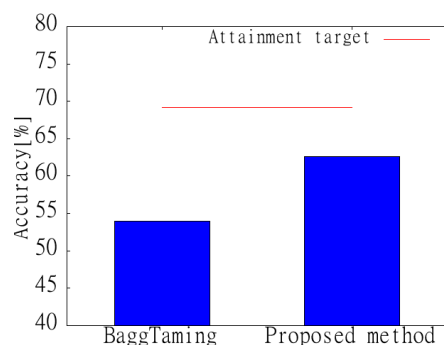


図 4.21: Distribution of original labels of “Abalone Dataset” in binary classification

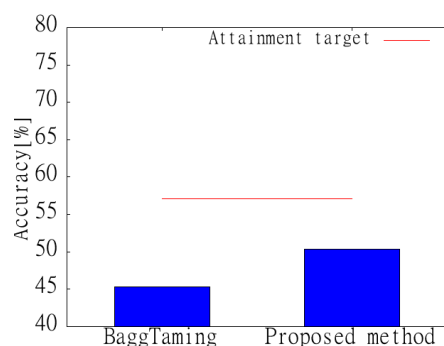


図 4.22: Distribution of original labels of “Abalone Dataset” in multilevel classification

場合の正判別率を到達目標として併せて示す。また，“Wine Quality Dataset”と“Abalone Dataset”の二値分類，多値分類のそれぞれを図 4.23，4.24，??，??に示す。表 4.21 から，すべてのラベルが正しい場合における正判別率は，“Wine Quality Dataset”の二値分類では 75[%] 程度，多値分類では 60[%] 程度となっていることがわかる。したがって，これらの値が本実験で使用したデータセットの到達目的であるといえる。また，図 4.23 から??より，従来手法では到達目標に対して差が大きいことがわかる。一方，提案手法では従来手法と比較して，到達目標に近づいていることがわかる。このことから，提案手法を用いることで，従来手法よりも到達目標に近い正判別率を示すことができ，データを有効に活用できていると言える。

表 4.21: Average accuracy of each method and attainment target

	Wine Quality Dataset		Abalone Dataset	
	binary	multilevel	binary	multilevel
BaggTaming	66.25	51.97	73.44	58.92
Proposed method	73.01	55.89	77.52	62.16
Attainment target	74.12	59.13	77.77	65.63

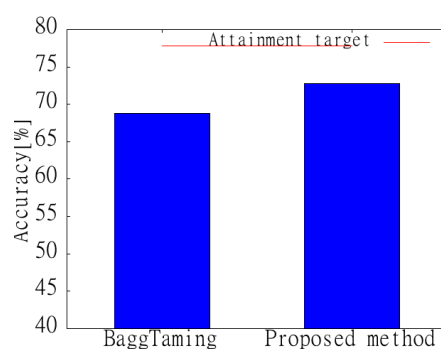


図 4.23: Average accuracy and attainment target using “Wine Quality Dataset” in binary classification

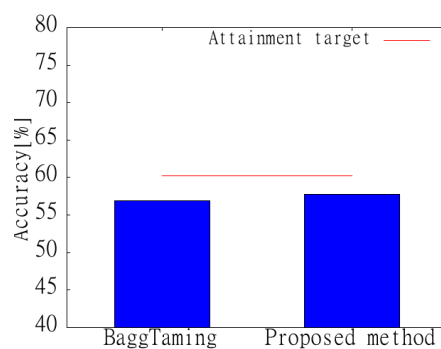


図 4.24: Average accuracy and attainment target using “Wine Quality Dataset” in multilevel classification

表 4.22: Contingency table

		Observation	
		Positive	Negative
Estimate	Positive	TP (True Positive)	FP (False Positive)
	Negative	FN (False Negative)	TN (True Negative)

また、本実験では、正判別率を用いて提案手法の精度検証を行った。機械学習における学習器のその他の尺度として、適合率、再現率、F 値などがある。適合率、再現率、F 値についてそれぞれ考察する。検証に用いるデータは、“abalone dataset” の飼育データ数 $N_T = 40$ を対象とする。

まず、それぞれの評価指標について説明する。表 4.22 に、二値分類における分割表を示す。表 4.22 を用いた場合、正判別率 (Accuracy), 適合率 (Precision), 再現率 (Recall) は以下の様に示すことができる。

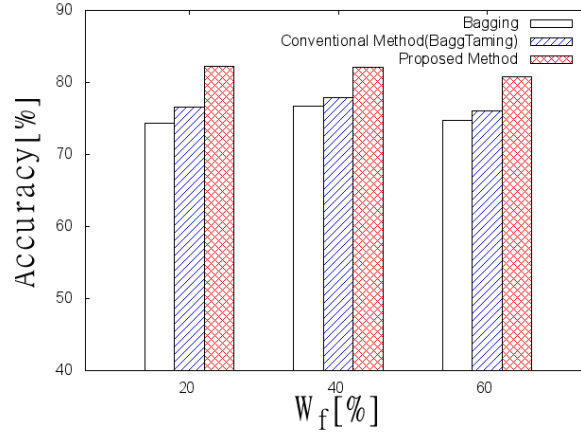
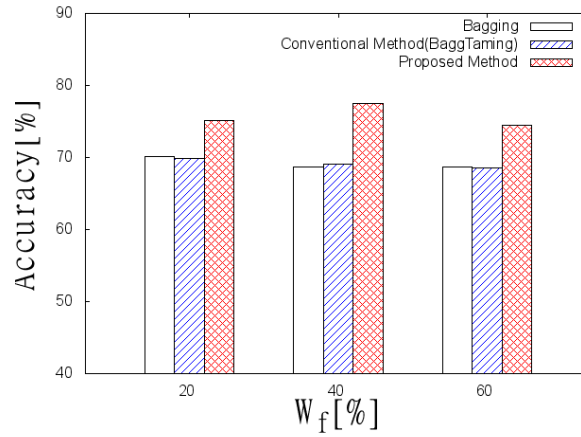
$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \quad (4.1)$$

$$Precision = \frac{TP}{TP + FP} \quad (4.2)$$

$$Recall = \frac{TP}{TP + FN} \quad (4.3)$$

上記の式からわかるように、正判別率は予測結果の正答率を表す。適合率は予測結果が正のうち、実際に正であった割合を表している。したがって、適合率は予測正に対する正確性を示す評価尺度である。再現率は、実際に正であるうち、予測結果が正とした割合である。したがって、再現率は正のサンプルに対する網羅性を示す評価尺度である。再現率と適合率はトレードオフになりやすく、どちらか一方を高めようとするともう一方は下がるため、学習器の活用目的に合わせた評価尺度の選定が必要である。また、F 値 (F-measure) は適合率と再現率の調和平均であり、以下の式で求めることができる。

$$F - measure = \frac{2RecallPrecision}{Recall + Precision} \quad (4.4)$$

図 4.25: Precision for tame data using “Abalone Dataset”, $N_T = 40$ 図 4.26: Recall for tame data using “Abalone Dataset”, $N_T = 40$

F 値を用いることで、適合率と再現率の2つの尺度をまとめることができ、学習器自体の性能評価が容易となる。

図 4.25 に Bagging, BaggTaming, 提案手法の適合率を示す。図 4.25 からわかるように、提案手法では Bagging や BaggTaming と比較してより高い値を示している。したがって、これらの手法と比べ、実際に正であるサンプルに対する正答率が高まったということがわかる。

図 4.26 に Bagging, BaggTaming, 提案手法の再現率を示す。これも適合率と

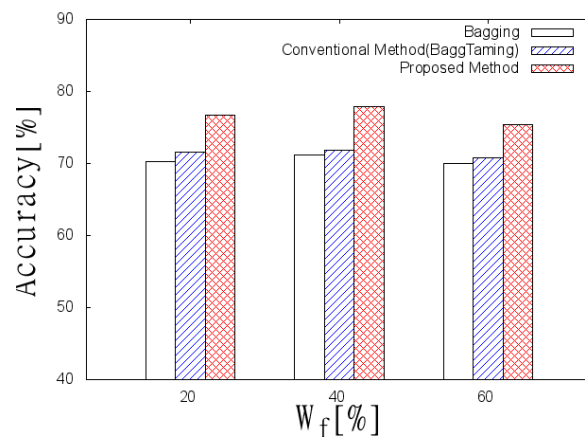


図 4.27: F-measure for tame data using “Abalone Dataset”, $N_T = 40$

同様に，提案手法が Bagging や BaggTaming と比較してより高い値を示している．したがって，より多くの正であるサンプルに対して正のラベルを付けることができ，網羅性が高いことがわかる．また，適合率，再現率の二つの指標において，ともに従来手法よりも向上していることがわかる．このことから，提案手法は FP, FN を共に低減させ，その減少分を TP に与える指標であることがうかがえる．図 4.27 に Bagging, BaggTaming, 提案手法の F 値を示す．適合率，再現率の結果からも推測できるように，提案手法の F 値は Bagging や BaggTaming よりも高い値を示している．したがって，正判別率以外の評価指標として適合率，再現率，F 値を用いた場合でも，提案手法は高い性能を示すといえる．

4.6 提案手法のまとめ

本章では，3 章で提案したアルゴリズムについて，汎用データセットを用いた効果検証を行った．意図的に誤ったラベルを持つ訓練事例を混入させた汎用データセットを用いて，提案手法の検証を行った．提案手法の検証項目として，ラベルの推定精度，誤った訓練事例数の推移，判別能力の三項目を検討した．ラベル推定精度は，飼育データ数 N_T が大きくなるほど向上することが分かった．誤っ

た訓練事例数の推移は、提案手法を適用することでいずれの条件でも誤ったラベルを持つ訓練事例を減少させることができ、また野生データに含まれる誤った訓練事例が多いほど効果が高いことが分かった。提案手法の判別能力は、飼育データのみを用いた Bagging よりも高く、従来手法である BaggTaming よりも高いことが分かった。この判別能力の向上は、野生データに含まれる誤った訓練事例に対して、学習目標に従って正しくラベルを与えることができたため、学習に活用可能な訓練事例数が増加したことに起因すると考えられる。また、提案手法について、正判別率以外の指標を用いた検証も同時に行った。検証では、新たな指標として適合率、再現率、F 値の三つの評価尺度を用いた。いずれの評価尺度を用いた場合でも、提案手法は飼育データのみを用いた Bagging よりも高く、従来手法である BaggTaming よりも高い値を示すことが分かった。

以上のことから、提案手法を用いることで、誤った訓練事例のラベルを適切に修正することができ、取得可能な限られたデータを有効に活用することができるといえる。

第5章 ワイヤボンディングにおける 抜取検査を題材とした 実証実験

5.1 はじめに

マイクロエレクトロニクス的发展により，電子機器の小型化，高機能化が急速に進んでいる．それに伴い，半導体実装技術も微細化，高密度化が進み，素子の接続数の増加から接続部における信頼性は一層重要度を増してきている．

ワイヤボンディングは半導体の接合技術の一つとして広く採用されている．近年ではフリップチップ接続 [94] や TAB(Tape Automated Bonding) [95] などのワイヤレスによる接続方式が広まりつつあるが，配線の自由度の高さや基盤技術が確立されていることによる低コスト化などの要因により，ワイヤボンディングは今なお広く活用されている．

ワイヤボンディングの品質保証には主に破壊試験を伴う抜取検査が用いられている．抜取検査では，「ロット」と呼ばれるいくつかの生産サンプルからなる集合を構成し，サンプリング計画に基づいてロットの一部のサンプルを抜き取り，品質試験を実施し合否を検査する．試験の結果をロット内にある全サンプルで共有することで，全サンプルに対する品質を決定している．本稿では，ロットから選択され品質試験を実施するサンプルを「検査サンプル」と呼び，その他のサンプルを「未検査サンプル」と呼ぶ．抜取検査は，品質試験に係るコストが高い製品や，破壊試験などによる品質保証のため原理的に全数検査が不可

能な製品などに対して統計的に品質を保証する検査法である。

しかし、抜取検査はロット内の多くのサンプルに対して品質試験を実施しないため、突発的に発生する不良に対応できないという課題がある。また、抜取検査によって不良とされたロットはすべて廃棄されてしまうため、良品の誤廃棄などによる歩留まりの低下も問題となる。特に、半導体は高密度化、高度化の進展により製造工程数が非常に多く製造に係るコストも高い。さらに、ワイヤボンディングは接合面の粗さや汚れの付着によって接合強度が低下する [96] など、接合ターゲットの状態に起因する接合強度のばらつきもあるため、同一ロット内のサンプルにおいても接合状態にばらつきが生じる。したがって、現行の抜取検査によるロット単位で行う品質保証ではなく、個々の生産サンプルに対する品質保証が求められている。

ワイヤボンディングの接合方式として熱圧着に加え超音波を印加するサーモソニック方式が広く採用されている。超音波エネルギーは接合部の塑性変形と相互拡散に利用されており、超音波を印加することで従来の熱圧着方式よりも強固な接合が可能となることが知られている。ワイヤボンディング工程における印加超音波の解析により、接合状態に関する有意な情報の取得可能性が示唆されている [97,98]。本研究では、ワイヤボンディング工程における印加超音波に着目し、印加超音波に基づいた全数に対する品質推定手法を提案する。

全サンプルに対して品質を評価する手法として教師あり学習を用いた品質推定法がある。一般的な教師あり学習では、個々の入力データに対して教師ラベルを与えた学習用データからモデルを学習させることで、所望する判定基準を教師ラベルに基づいて獲得することができる。ワイヤボンディングへの適用としては、入力データはワイヤボンディングの印加超音波であり、教師ラベルは破壊試験の結果となる。教師あり学習では学習用データをあらかじめ取得しておき、それから学習したモデルを現場へインストールするという活用法が一般的であ

る．しかし、半導体は製品の開発サイクルが非常に短く更新が早いため、学習用データをあらかじめ取得する方法ではランニングコストを高めてしまう．そこで、抜取検査による品質保証と機械学習による品質推定を併用する品質推定法を提案する．生産初期は抜取検査によって品質を保証するとともに学習用データを蓄積し、定常生産期には抜取検査による品質推定と蓄積した学習用データから学習した判別モデルによる全数に対する品質推定を行う．本研究で提案する品質推定手法は、定常生産期には抜取検査ではできなかった全数に対する品質評価が実現でき、さらに機械学習に用いる学習用データの事前収集を必要としないため、機械学習のインシヤルコストの抑制につながる．

本研究では抜取検査の枠組みを活用し、飼いならし学習を導入する．飼いならし学習では、信頼性の高い教師ラベルを持つ少数の飼育データと、信頼性の低い教師ラベルを持つ多数の野生データを併用する機械学習である．抜取検査に当てはめた場合、高信頼かつ少数のデータである飼育データは、センサ情報と品質ラベルが明確に紐づいた「検査データ」である．また、低信頼かつ大量のデータである野生データは、センサ情報と品質ラベルが明確に紐づいたものと、そうではないものが混在している「未検査データ」である．飼いならし学習の従来手法である BaggingTaming では、野生データから飼育データと同水準で教師ラベルを与えられたデータを選択的に学習に用いることで学習用データ数を補完し、飼育データ単体からの学習よりも高精度な学習を行うことを目的としている．一般的な教師あり学習では教師ラベルはすべて信頼できることを前提としているため「検査データ」のみを学習に用いるが、飼いならし学習では「検査データ」とともに「未検査データ」も活用するため、効率的に学習用データを収集することができる．

一般に野生データにはいくつかのデータが飼育データと同水準でラベルを与えられているかは未知であるため、学習用データとしていくつかの野生データが活

用できるかは不明である．そこで，提案手法を用いて野生データ中にある誤ったラベルを持つ訓練事例を修正することで，より多くの訓練事例を活用した学習が可能となり，「未検査サンプル」につけられた品質ラベルの信頼性が低い場合でも，効果的に学習することが期待できる．

本章では，提案手法を用いたワイヤボンディング抜取検査の性能向上法を提案する．本研究の目的は，抜取検査における枠組みを活用した品質推定を実現し，破壊試験に係るコストを増加させずに全数に対して品質を評価することで，抜取検査の信頼性向上に寄与することである．実際に作製したボンディングサンプルを用いて提案手法を適用し，その判別能力を抜取検査と従来手法との比較を通じて検証する．

5.2 機械学習のワイヤボンディング抜取検査への適用

5.2.1 ワイヤボンディング

ワイヤボンディングはIC製造工程において重要な工程の一つであり，半導体チップと外部リードとの接続技術である．ワイヤボンディングの特徴として，自由度の高い接続が可能であるため，多様な対象に接続可能であること，基盤技術が確立されているため管理コストを低くできるなどが挙げられる．ワイヤボンディングの接合例を図5.1に示す．近年では，ファインピッチ化に対応した技術開発により，さらなる微細化が進むと考えられている．さらに，近年では電子機器の小型化，高機能化が急速に進んでおり，接続素子の数が膨大となってきたため，接続部の信頼性はより一層重要となってくる．

ワイヤボンディングの接続は従来の熱圧着に加え超音波を印加するサーモソニック方式が広く利用されている．サーモソニック方式により， $150^{\circ}\text{C}\sim 200^{\circ}\text{C}$ 程度の比較的低温での接合が可能となり，材料の変成が少なくなる．ワイヤボンディングの十分な接合には温度管理と塑性変形による新生面の生成が不可欠

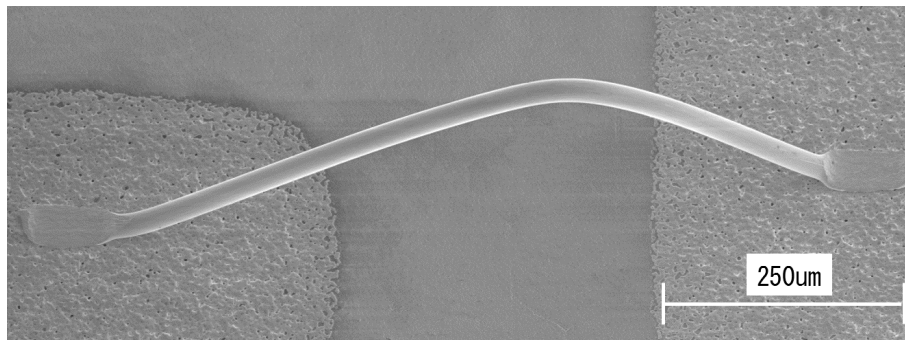


図 5.1: wire bonding

であることが知られており [99], 超音波エネルギーは塑性流動とともに相互拡散を促進させるために利用される. また, ワイヤボンディングの接合部の強度は印加した超音波の振幅および周波数の増大によって上昇することが報告されている [100]. したがって, ワイヤボンディングにおいて印加超音波の振幅と周波数は接合強度向上のために重要な因子となる.

5.2.2 ワイヤボンディングの信頼性評価とその課題

ワイヤボンディングの品質評価法として, ワイヤプルテストなどの破壊試験が一般的に用いられている. 破壊試験は, 外力を加えることで破断や亀裂の発生やその成長過程を観察することで, その製品の良否を調べる試験方法である. ワイヤプルテストでは, ワイヤに鉛直上向きの荷重を加えてワイヤを破断する. ワイヤ破断時の荷重と破断モードと呼ばれるワイヤの破断位置を記録し, それらから統合的に判断して製品の良否を決定する. ワイヤプルテストの概要図を図 5.2 に示し, プルテストに用いるプルテストタを図 5.3 に示す.

破壊試験はその性質上, 生産ラインにおけるすべての製品に対して実施することができない. そのため, 実際の生産ラインでは, 破壊試験を伴う抜取検査による統計的な品質保証を行うにとどまっている. 抜取検査は, 検査対象の集

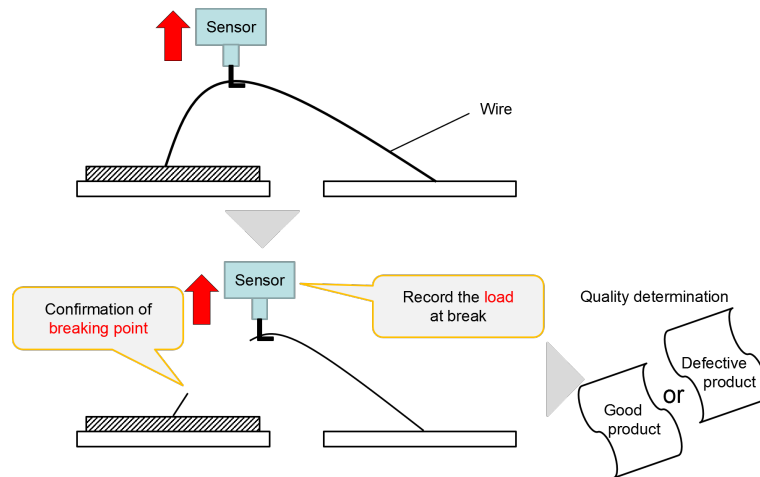


図 5.2: Overview of pull test

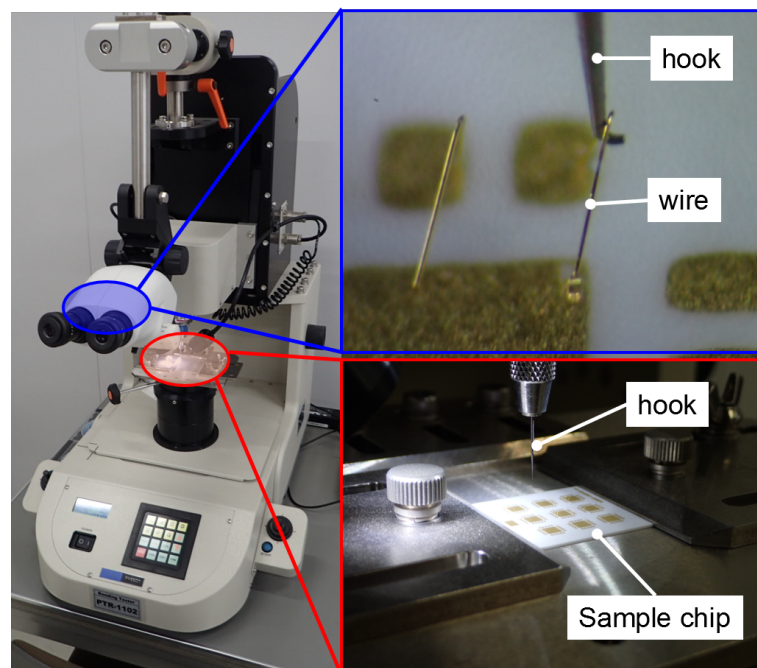


図 5.3: pull tester

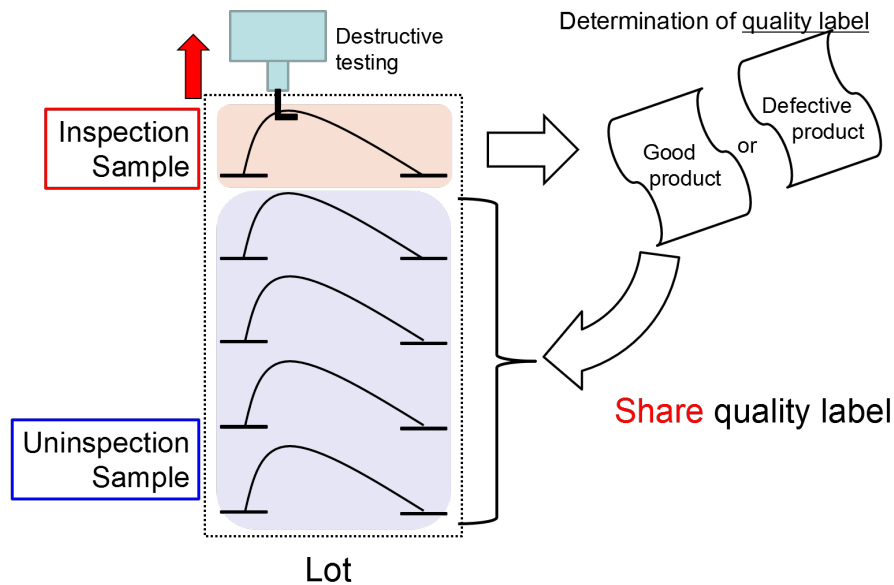


図 5.4: Overview of Sampling inspection

合であるロットに対して、その一部をあらかじめ定められた方式で選択、試験を行い、その試験結果に基づいてロット全体の合否を判定する検査手法である。抜取検査の具体的な手順としては、以下の様になる。

1. サンプルング計画に基づいて一定数の検査対象を集め、ロットを構成する。
2. 構成したロットからあらかじめ定められた方式で代表サンプルを選択し、破壊試験による品質評価を行う。
3. 代表サンプルの試験結果に基づき、ロット全体の合否を判定する。

抜取検査の概要図を図 5.4 に示す。

ワイヤボンディングでは、破壊試験によって得られるワイヤ破断時の耐荷重のほかに、破断モードを考慮した統合的な品質評価がなされている。しかし、一般的なワイヤボンディングの破壊試験では、同じ耐荷重でも異なる破断モードが含まれる、接合部が小さいためループの形状や負荷位置のわずかなずれが耐

荷重や破断モードに敏感に影響を与えるなど、破壊試験の結果が必ずしもワイヤボンディングの接合状態を評価しているとは言えないという問題が指摘されている [101]。また、ワイヤボンディングなどの半導体製品は技術の進展が早く、パッケージの形態やパッドピッチは日々変化している。そのため、ベンチマークと呼ぶべきデータセットの構築が困難であり、接合と品質評価に関する知見の蓄積がうまく進まないという現状がある。さらに、抜取検査では、代表サンプルの試験結果によってロット全体の合否を判定するため、ロット内に含まれるサンプルは、統計的なばらつきを少なくする必要がある。ロットを構成する検査対象が少ないほど、抜取検査の信頼性は高くなるが、破壊試験を伴う抜取検査では、代表サンプルは製品としての機能を失う。そのため、原理的に全数検査への移行は不可能であり、破壊試験を伴う抜取検査は品質の確保とコストにおいて常にトレードオフの関係にある。また、抜取検査は統計的な品質確保にとどまるため、突発的に発生した不良の流出や良品の誤廃棄に対応できないという課題もある。

したがって、破壊試験、抜取検査の両方から以下の課題が挙げられ、ワイヤボンディングの品質判定性能向上のために、全数に対するワイヤボンディングの接合状態に基づいた品質決定法が望まれている。

1. ワイヤボンディングの接合状態に基づいた品質決定
2. 検査コストを抑えつつ全数に対する品質決定

5.2.3 薄型 A E センサを用いた印加超音波の計測

本研究では、ワイヤボンディング時に印加される超音波に着目し、接合ターゲットを通して伝播する弾性波の計測技術を導入する。ワイヤボンディングの接合には新生面の生成が不可欠であり、印加される超音波は新生面の生成の促

進に用いられている。したがって、印加超音波には接合状態に関する有意な情報が含まれていると考えられ、これを計測・解析することで接合状態の推定が期待できる。しかし、接合ターゲットの表面にボイドや不純物がある場合や、接合時の圧力不足などによって十分に超音波が伝播しないなど、様々な要因で接合不良が発生することが考えられる。そのような接合不良の有無を含めた接合強度を評価するために、印加超音波を計測するのではなく、接合ターゲットを介して伝播する弾性波を計測することで、実際に接合部に印加された超音波を計測することができる。この弾性波から接合状態の推定を試みる。

弾性波の計測には、一般的にA E (Acoustic Emission) センサが用いられている。しかし、一般的なA E センサは80℃以上の高温環境での使用はできないため、150℃から200℃の高温環境下で接合するワイヤボンディングでは直接計測することができない。ウェーブガイドを用いることで弾性波の計測は可能となるが、弾性波の伝播中に微小信号が減衰してしまい、有意な情報の取得ができなくなる恐れがある。したがって、弾性波の計測は接合部直近に配置したセンサによる直接計測が望ましい。

本研究では、薄型A E センサを用いた弾性波計測技術を導入する。薄型A E センサは、国立研究開発法人 産業技術総合研究所で開発されたA E センサである。薄型A E センサの外観図を図5.5に示す。薄型A E センサの特徴として、800℃以上の過酷環境下においても計測可能であるという高い耐熱性が挙げられる。そのため、従来のA E センサでは直接計測不可能であった高温環境下でも安定動作可能となる。また、厚さが約1mm、1辺が5mm程度と従来のA E センサと比較して非常に小さい。半導体は微小な製品であるため、ワイヤボンディングの製造は狭小領域で行われることが多い。薄型A E センサは従来のA E センサと比較して狭小領域への取り付けの面でもアドバンテージを持つ。この二つの要因により、薄型A E センサを用いることで高温かつ狭小領域であるワイ

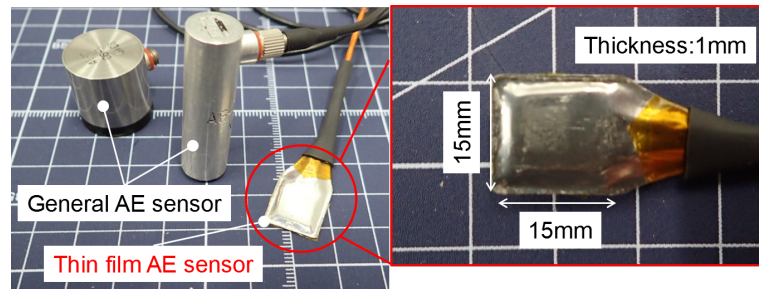


図 5.5: General AE sensor and thin film AE sensor

ヤボンディング接合部直近において、初めて直接計測可能となる。薄型AEセンサを用いて取得した弾性波を解析することで、接合状態に関する有意な情報の取得を試みる。

5.2.4 弾性波の解析と機械学習に用いる特徴量

ボンディング時に印加した弾性波をAEセンサを用いて計測する。AEセンサを用いて取得した弾性波は一般にAE信号と呼ばれる。ワイヤボンディングはIC素子上の電極とリード電極間を接合する技術であるため、一つのボンディングワイヤに対して二つの接合部（1st ボンディング，2nd ボンディング）を持つ。図5.6に計測したワイヤボンディング接合時のAE信号取得の例を示す。計測したAE信号から、接合強度に関する特徴量を算出する。まず、取得したAE信号から、接合時の信号を抽出する。抽出したボンディング接合時の信号から、機械学習に用いる特徴量をそれぞれ算出する。本実験では取得した信号に対して、時間領域と周波数領域からそれぞれ特徴量を算出した。周波数領域からは、フーリエ変換によって得られるスペクトルを8領域に等分し、それぞれの領域から平均、分散を算出し、計16次元を取得した。周波数領域から取得する特徴量の概略を図5.7に示す。時間領域からは、基本統計量（平均、分散、歪度、尖度）と、AE法に基づくAEパラメータ [102]（最大振幅、立上り時間、持続時間、AEカウント、立上り勾配、RMS）の計10次元を取得する。AEパラメー

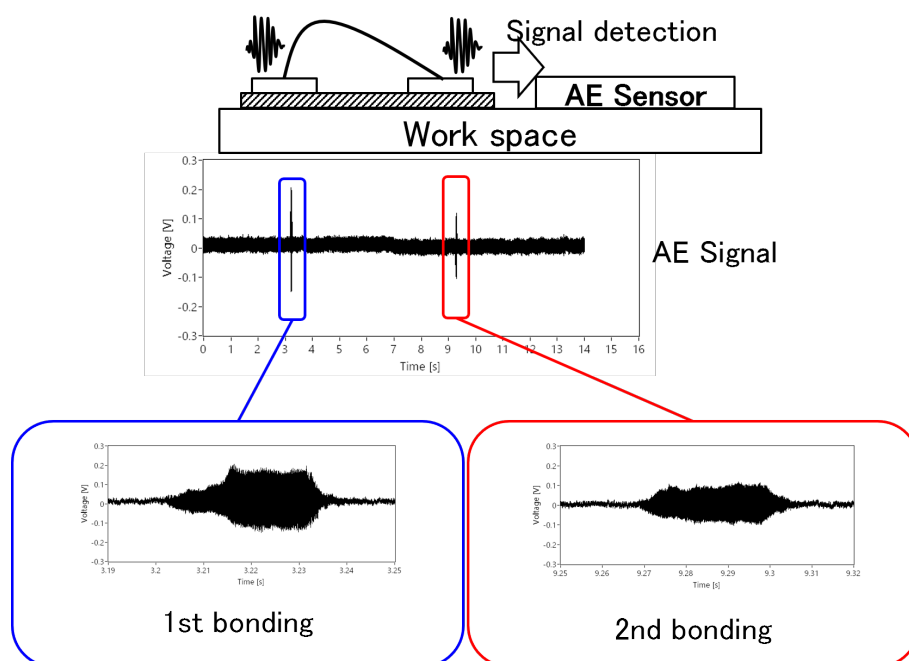


図 5.6: AE signal

タの概略図を図 5.8 に示す。

ボンディングサンプルには、接合部が二つあるため、1st ボンディング、2nd ボンディングについてそれぞれ時間領域特徴量と周波数領域特徴量を取得する。したがって、1 サンプルのボンディングワイヤから、時間領域特徴量 20 次元 (10 次元 \times 2)、周波数領域特徴量 32 次元 (16 次元 \times 2) の特徴量を取得することになる。

5.2.5 提案手法を用いた品質推定

取得した弾性波から得られる特徴量を入力とした機械学習を適用することで、全数に対する接合状態の推定を行う。前述のように、半導体分野においてパッケージング技術は急速に発展しており、大量のデータを確保することは高コストである。さらに、ワイヤボンディングの代表的な評価法である破壊試験では、

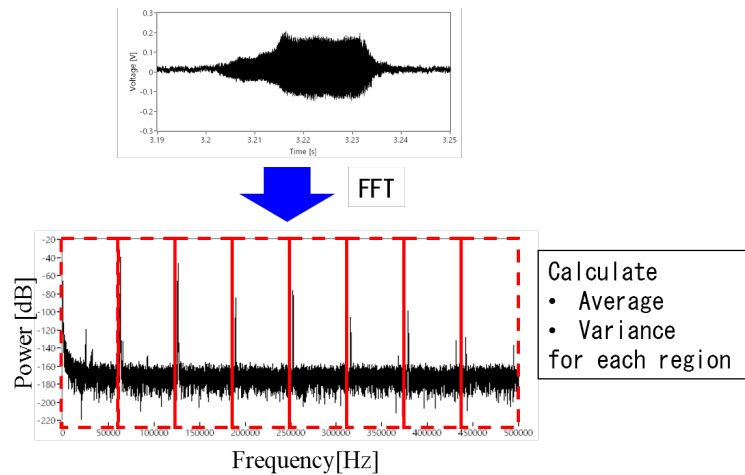


図 5.7: Feature values in the frequency domain

ループ形状によって大きく結果が異なってしまう。パッケージの形状や接合素子間の位置ずれによってワイヤボンディングの適切な形状も異なってくるため、異なるループ形状を用いた場合に適切な推定は困難であると考えられる。したがって、実用的な品質推定のために、現行の生産ラインで取得可能な限られた訓練事例集合を活用することが望ましい。

本研究では、現行の生産ラインで利用されている抜取検査の枠組みを活用したワイヤボンディングの品質推定手法を提案する。抜取検査では、ロット内に含まれるサンプルは、実際に破壊試験を行った「検査サンプル」と、破壊検査を行っていない「未検査サンプル」に分類できる。「検査サンプル」には実際の破壊試験結果に基づいた品質ラベルを与えられている。一方、「未検査サンプル」は「検査サンプル」の試験結果から判定された品質ラベルを与えられている。「未検査サンプル」には、抜取検査による統計的な保証はなされているものの、個々のサンプルに対して真に与えられるべき品質ラベルがつけられているとは限らない。したがって、「未検査サンプル」は品質ラベルが正しくつけられている場合と、そうではない場合が混在しており、すべてのサンプルを完全に信頼することはできない。また、抜取検査において、「検査サンプル」は少数であり、「未

検査サンプル」は比較的多数である。以上の条件から、「検査サンプル」は少数だが品質ラベルが厳密に管理されている飼育データであり、「未検査サンプル」は多数だが品質ラベルが厳密に管理されていない野生データであるといえる。

通常の機械学習では完全に信頼できる「検査サンプル」のみを用いて学習を行うが、抜取検査の特性上、多くのサンプルを集めることは困難となる。これに対して飼いならし学習では、「検査サンプル」と「未検査サンプル」を併用する。「未検査サンプル」の中から利用可能なものを選択し、学習に活用できるため、「検査サンプル」のみを用いた場合よりも高い品質推定能力を示すと考えられる。本研究では、さらなる「未検査サンプル」の利活用を促すため、提案手法による「未検査サンプル」の品質ラベルの修正を試みる。「未検査サンプル」の品質ラベルを修正することで、従来手法よりも多くの「未検査サンプル」の学習の活用と、従来手法よりも高い品質推定性能の実現が期待できる。

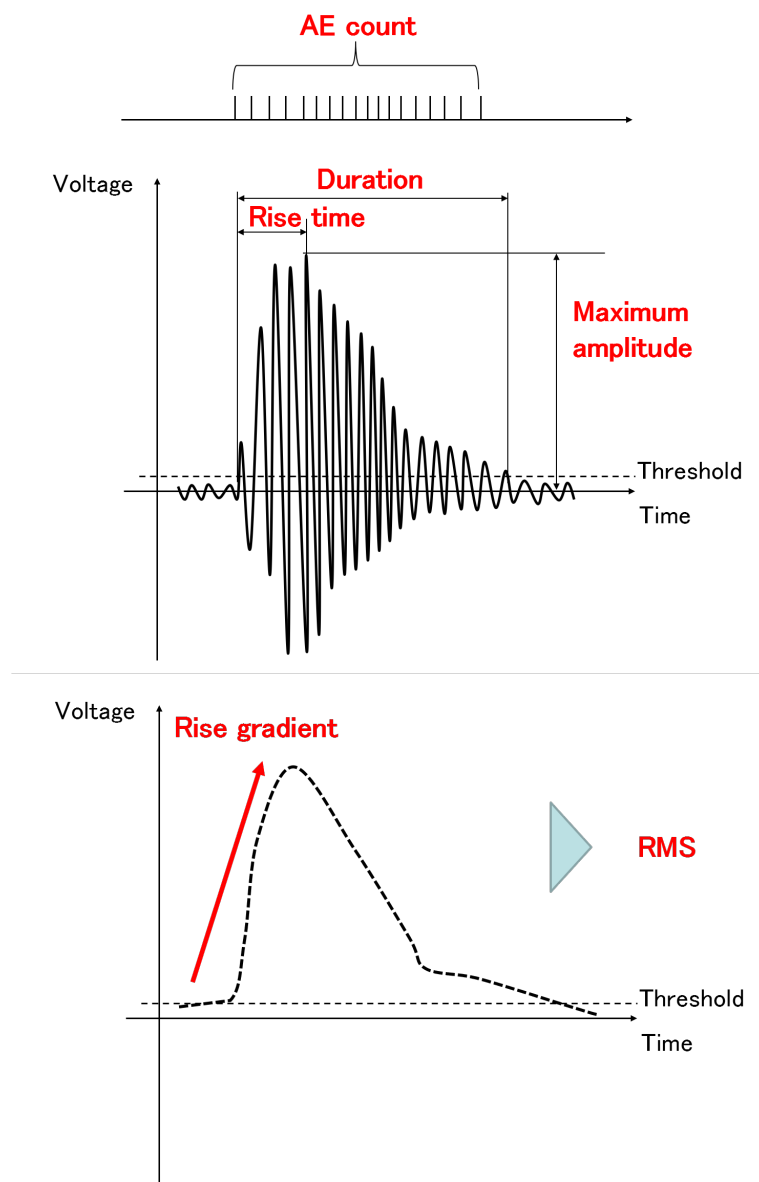


図 5.8: AE parameter

表 5.1: Bonding parameter

Ultrasonic power	360[kHz]
Ultrasonic application time	50[ms]
Heater temperature	210[°C]
load	50[g]

5.3 実験

5.3.1 実験環境

提案手法による品質ラベル推定精度の検証を行う。本実験では、金メッキされたサンプルチップを用いてボンディングサンプルを作製した。ボンディングワイヤには Au ワイヤ（直径 $25.4 \mu\text{m}$ ）を使用した。接合に用いる機器として、卓上型マニュアルワイヤボンダ（モデル 7400 D，WEST BOND 社）を用いる。ワイヤボンディングの接合環境として、ヒータに接続されたワークホルダ上にサンプルチップと薄型 A E センサを図 5.9 に示すように配置した。これにより、従来の A E センサではできなかった接合部近傍での A E 信号の計測を行う。また、ヒータの設定温度は 210°C とした。計測した A E 信号は、差動増幅器を用いて増幅し、アナログ入力モジュール（NI9223，National Instruments 社）を用いて PC へ保存した。アナログ入力モジュールのサンプリングレートは 1 M Sample/sec とした。実験環境の外観を図 5.10 に示す。本実験で行うワイヤボンディングに関する各種パラメータについて表 5.1 に示す。

5.3.2 データセットの構成

本実験では、学習用のサンプルとして、1770 サンプルのボンディングワイヤを作製した。これらのサンプルから抜取検査を模擬してロットを構成し、学習に用いる。抜取検査では、サンプリング計画によってさまざまなデータセットの構成が考えられるが、本実験では単純なサンプリング計画を採用した。まず、

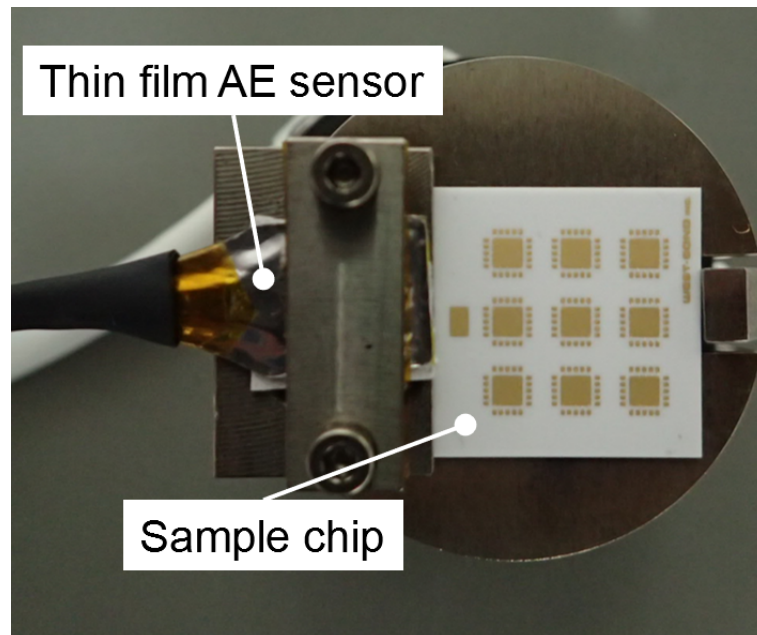


図 5.9: AE Sensor installation

構成するロットサイズは5とした．すなわち，各ロットに含まれるボンディングワイヤは5サンプルとなる．また，各ロットから抜取検査に用いる検査サンプルは1サンプルとした．したがって，各ロットの内訳は検査サンプル：1サンプル，未検査サンプル：4サンプルとなる．このロットをすべての学習用サンプルを用いて構成した．ロット数は全部で354となる．

次に，抜取検査の枠組みに従い，各ロットの検査サンプルに対して破壊試験を行い，その結果に基づいてロットごとに良否を判定する．本実験では破壊試験としてワイヤプルテストを採用した．破壊試験に基づいた良否判定の結果，良品ロットは110，不良品ロットは224となった．したがって，全1770サンプルのうち，良品サンプルは550サンプル，不良品サンプルは1220サンプルとなる．

また，本実験では，5分割の交差検証による評価を行う．本実験では良品サンプルと不良品サンプルの数の比はおよそ1対2となるため，良品：110サンプル，不良品：224サンプルの計334サンプルを評価用サンプルとして用いた．この評

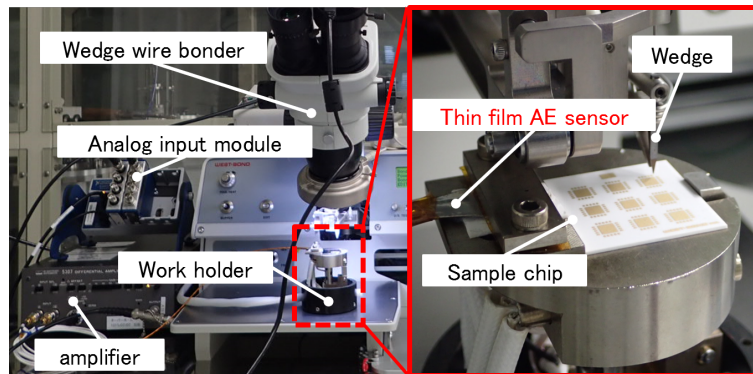


図 5.10: Experiment environment

表 5.2: Number of samples used for experiment

Lots	Inspection samples	Uninspection samples
40	40	160
60	60	240
80	80	320
100	100	400

価用サンプルに対する判別能力を用いて，提案手法の有効性を検討する．

5.3.3 実験条件

前節で構築したデータセットに対して，提案手法による品質ラベル推定を行う．まず，学習に用いるサンプルを，データセットからランダムに抽出する．本実験では，学習に用いるロット数 $N = \{40, 60, 80, 100\}$ と変動させながら実験を行った．本実験で準備したデータセットは，良品ロット数と不良品ロット数が一致していないが，学習に用いる良品ロット数と不良品ロット数は一致するように選択した．ロット数と検査サンプル数，未検査サンプル数の組合せは表 5.2 のようになる．本実験では抜取検査を模擬しているため，従来手法と提案手法で使用される未検査サンプルは，選択されたロットに応じて一意に決定した．したがって，学習に用いるロットの数によって未検査サンプル数も変動し，また従

来手法と提案手法で使用する未検査サンプルは一致している。

実験は各条件において10回ずつ行った。それぞれの手法を用いた場合の判別能力の評価指標として、テストサンプルに対する判別性能を用いた。また、従来手法、提案手法ともに、学習器としてアンサンブル学習で一般的に使用される決定木を用いた。従来手法と提案手法の判別器の採用基準として、検査サンプルから学習した学習器の、検査サンプルに対する判別率を用いた。

5.3.4 実験結果・考察

構築したデータセットを用いた良否判定を行う。本実験では、提案手法を用いた良否判定の比較対象として、検査サンプルを用いて学習を行った Bagging と、検査サンプルと未検査サンプルを用いた BaggTaming の結果を併せて示す。

まず、時間特徴量を用いた場合について考察する。図 5.11 に、時間特徴量を用いた場合のテストサンプルに対する判別率を示す。また、詳細な値を表 5.3 に示す。図 5.11 から、検査サンプルのみを用いた Bagging の判別率は、検査サンプルすべての条件において低くなっていることがわかる。これは、学習に用いたサンプルが少ないことに起因すると考えられる。一般的な機械学習では、十分な判別能力を示すためには、特徴量の次元数よりも十分に多いサンプルから学習を行う必要があるということが知られている。本実験では、 $N = 100$ の条件でも次元数の5倍のサンプルしか学習に用いることができないため、判別能力が低くなったと考えられる。

また、従来手法である BaggTaming の判別率は、検査サンプルのみを用いた Bagging よりも高くなっていることがわかる。これは、BaggTaming の手続きにより、未検査サンプルから検査サンプルと同じ基準で品質ラベルを与えられたサンプルを学習用サンプルとして取り込むことができたためであると考えられる。未検査サンプルから活用可能なサンプルを学習に取り込むことで、検査サ

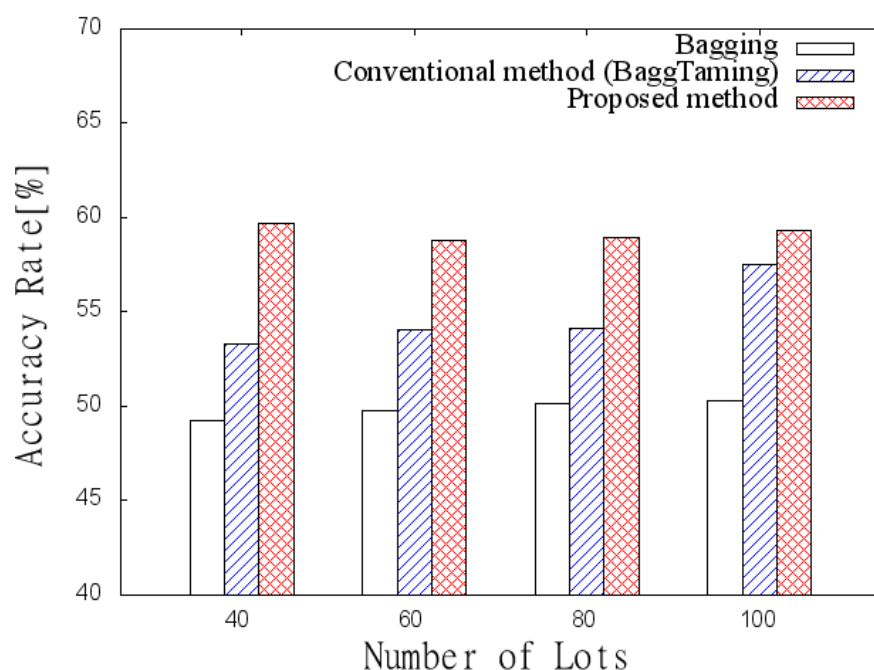


図 5.11: Accuracy using Time features

サンプルのみを用いた場合よりも多くのサンプルから学習ができ、判別率の向上につながったといえる。BaggTaming の判別率と Bagging の判別率の比較から、機械学習では多くのサンプルを用いた学習が判別能力の向上に必要であることが再確認できる。

提案手法では、BaggTaming よりもさらに高い判別率を示している。これは、提案手法が有効に作用したためであると考えられる。提案手法では、BaggTaming の手続きにおいて学習用として取り込み不可とされたサンプルに対して、与えられるべき品質ラベルを推定・付与したのち、再度学習用サンプルとしての取り込み判定を行っている。したがって、提案手法は BaggTaming よりも多くのサンプルから学習することができ、提案手法の判別率は BaggTaming よりも高くなったと考えられる。

図 5.12 に、周波数特徴量を用いた場合のテストサンプルに対する判別率を示

表 5.3: Accuracy rate using Time features

Lots	Bagging	BaggTaming	Proposed method
40	49.27	53.31	59.72
60	49.75	54.08	58.81
80	50.17	54.12	58.97
100	50.33	57.49	59.31

表 5.4: Accuracy rate using Frequency features

Lots	Bagging	BaggTaming	Proposed method
40	49.29	53.66	60.39
60	50.38	50.17	59.02
80	49.44	50.77	59.05
100	49.52	53.63	58.31

す。また、詳細な値を表 5.4 に示す。この場合も時間特徴量と同様に、検査サンプルのみを用いた Bagging の判別率は、すべての条件において低くなっていることがわかる。従来手法である BaggTaming の判別率は、検査サンプルのみを用いた Bagging よりも高くなっており、提案手法は BaggTaming よりもさらに高い判別率を示すことがわかる。周波数特徴量の結果と時間特徴量の結果を比較すると、周波数特徴量を用いた場合の提案手法と BaggTaming の判別率の差は、時間特徴量を用いた場合よりも大きい。このことから、時間特徴量が周波数特徴量よりも少ない次元数であることと、A E 法に基づいた検査によって有意な情報が獲得できたことがうかがえる。

図 5.13 に、時間特徴量と周波数特徴量を併用した場合のテストサンプルに対する判別率を示す。また、詳細な値を表 5.5 に示す。この場合も、時間特徴量や周波数特徴量の場合と同様に、検査サンプルのみを用いた Bagging は判別率が低く、次いで従来手法、提案手法と判別率が高くなる。

また、本実験の判別率は、ほとんどの実験で 70 % を下回る低い判別率を示した。本実験では、機械学習の入力として A E 信号から取得した特徴量を用いて

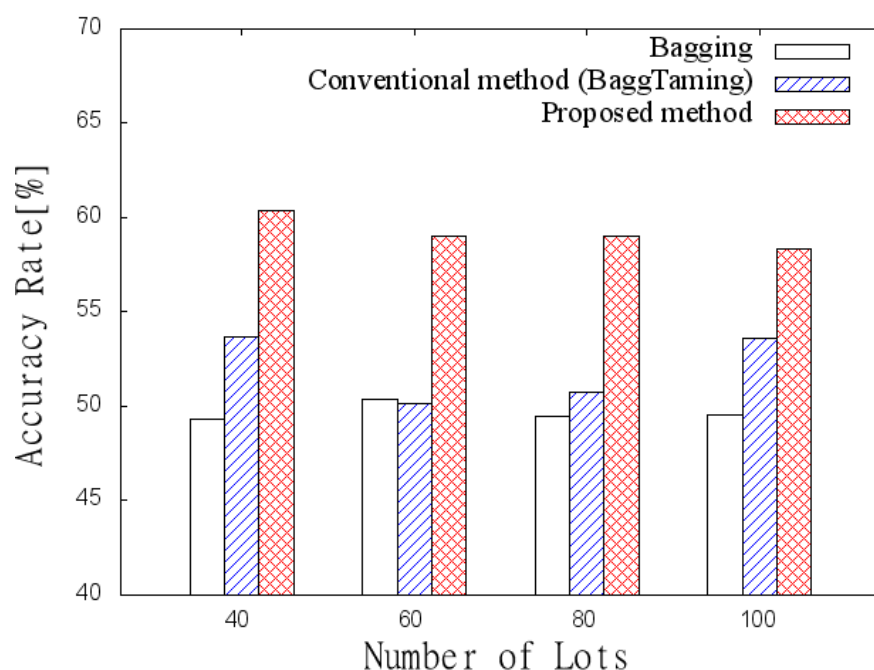


図 5.12: Accuracy using Frequency features

表 5.5: Accuracy rate using Time features and Frequency features

Lots	Bagging	BaggTaming	Proposed method
40	48.74	53.51	61.32
60	49.78	51.23	60.01
80	48.48	52.63	58.94
100	48.80	56.06	59.97

おり、出力としてワイヤプルテストによる良否判定結果を用いている。また、実験に使用したボンディングサンプルはマニュアルワイヤボンダを用いて作製している。そのため、ワイヤループの形状に起因するワイヤプルテストの評価の揺らぎがあったと考えられる。ワイヤループの形状は、A E 信号には表れない情報であるため、入力と出力の間にループ形状に起因する隠れ因子があり、それによって判別率が低くなったと考えられる。サンプルとして用いるワイヤループの形状を一定にすることで、ループ形状のばらつきに起因する評価の揺らぎを低減することができ、さらに高精度な品質判定が可能となると考えられる。

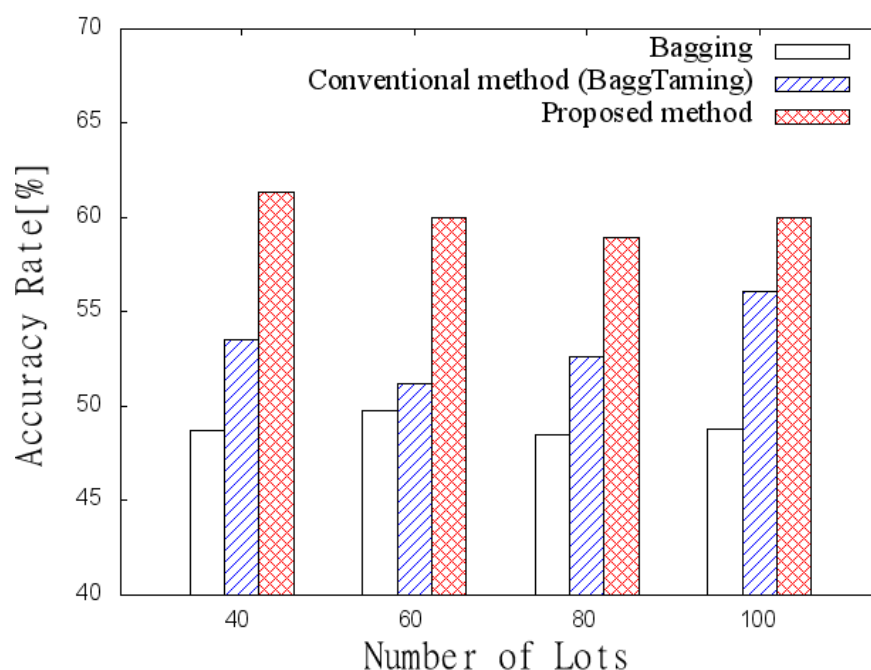


図 5.13: Accuracy using Time features and Frequency features

ロット数を大きくした場合の判別能力について、考察を行う。考察では、時間特徴量を対象とした。図5.14にロット数の変化に伴う各手法の正判別率の推移を示す。図5.14から、飼育データのみを用いた Bagging と従来手法の BaggTaming は、ロット数が増加するに連れて判別能力が高くなっていることがわかる。これは、ロット数の増加に伴い、学習に用いることができる検査サンプル数が増えたため、学習精度が向上したと考えられる。このことから、学習に用いるデータ数によって判別能力は変動し、少ない場合では高い性能を示すことができないということがわかる。また、従来手法の BaggTaming は、Bagging よりも急な勾配で判別能力が上昇していることがわかる。これは、ロット数の増加に伴う未検査データの増加数は、検査データ数よりも高いためであると考えられる。今回の実験では、ロットサイズは5とし、検査サンプル数を1としている。したがって、各ロットに含まれる検査サンプル数は1なのに対して、未検査サンプル

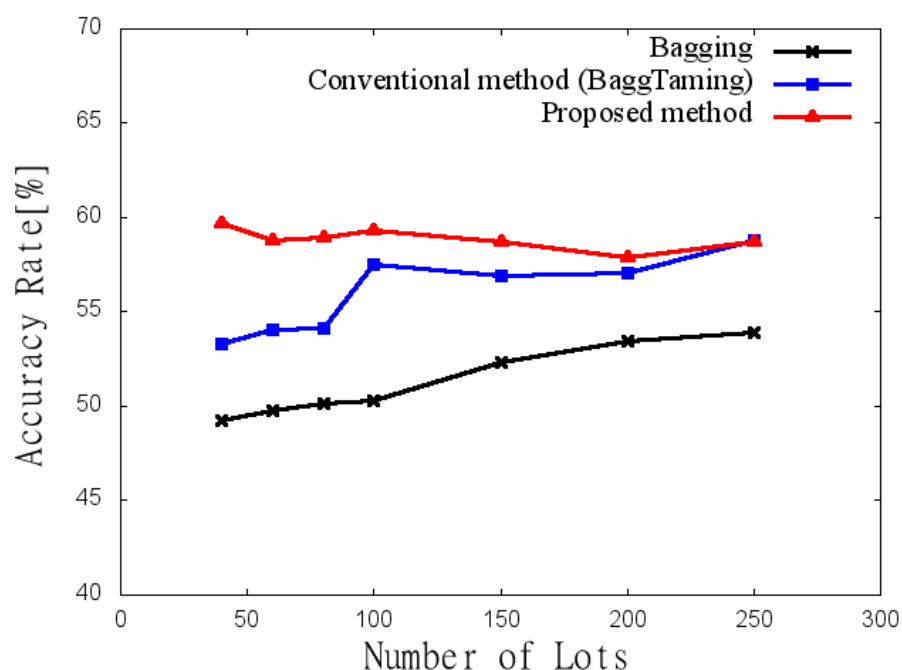


図 5.14: Discussion to accuracy using Time features

は4サンプル含まれる．これらを活用することができる従来手法の Baggtaming は，Bagging よりも多くのデータを活用することができるため，勾配が急になったと考えられる．

さらに，提案手法はロット数 40 の時点で 60[%] に近い判別能力を示し，以後ほぼ横ばいに推移している．これは，本サンプルを用いた場合の正判別率の上界が 60[%] 近傍であり，提案手法はロット数 40 の時点ですでに到達していると考えられる．このことから，提案手法はロット数が少ない場合でも，未検査サンプルから従来手法よりも多くのデータを学習に取り込むことができるということを示している．また，実環境への適用を考慮する場合，ロット数が少ない時は生産初期である．提案手法は，まだ検査サンプルが十分に確保できない生産初期において高い判別能力を示すことができ，本実験の場合では検査サンプルを用いて学習を行う Bagging と比較しておよそ 11[%] の判別能力の向上を示

している。以上のことから、提案手法は生産現場における生産初期段階において高精度な判別を実現できる手法であり、リアルデータの活用サイクルを早めるという点で非常に有効な手法であるといえる。

5.4 結論

本章では、薄型AEセンサによる弾性波の計測と提案手法を用いたワイヤボンディング抜取検査の性能向上法を提案した。マニュアルワイヤボンダを用いて作製した実サンプルを用いて抜取検査を模擬したデータセットを構築し、提案手法による品質推定を行った。本実験から、以下の結果が得られた。

1. 「検査サンプル」のみを用いて学習を行った場合、テストサンプルに対する品質推定性能は低くなっている。この原因は、「検査サンプル」が少なく、十分な学習に必要なデータ数が確保できないためである。
2. 「検査サンプル」と「未検査サンプル」を併用して学習を行う BaggTaming と提案手法は、「検査サンプル」のみを用いて学習した場合よりも有意に良い判別能力を示した。このことから、「未検査サンプル」から「検査サンプル」と同水準でラベルを与えられた訓練事例を適切に抽出することで、学習に用いる訓練事例数を補完し、結果的に「検査サンプル」を用いた場合よりも多くの訓練事例を活用して学習することができたといえる。
3. 提案手法は BaggTaming よりも高い判別能力を示す。この原因は、BaggTaming で取り込み不可とされたデータに対して、与えられるべき品質ラベルを適切に推定し、活用可能な情報へと昇華でき、それらの訓練事例を学習に活用できたためであると考えられる。

また、本研究で用いたボンディングサンプルはマニュアルワイヤボンダで作製したため、ワイヤループの形状に起因するワイヤプルテストの評価の揺らぎ

があったと考えられる．その結果として，評価用のテストサンプルに対する判別率は，最大で 70%程度と低い判別率となった．今後は，弾性波と品質検査の関係を厳密に評価するために，ワイヤボンディング工程の自動化によるワイヤループ形状の安定化を図る予定である．

第6章 結論

本研究は、第4次産業革命の流れを受け、生産現場における多様化する消費者ニーズへの対応と生産性・設備稼働率の向上に寄与するためのデータ利活用を促す新しいアルゴリズム開発を目的として、既存の生産現場における高信頼データの収集・管理コストに関する課題の解決および限られたデータの利活用アルゴリズムを目標として遂行された。以下に本研究で得られた知見を示す。

第2章では、本研究で扱う飼いならし学習問題の問題設定と扱う訓練事例集合について解説した。また、従来まで提案されてきた機械学習手法のうち、少数の高信頼データが取得可能な状況を扱う手法を説明し、それらとの飼いならし学習の違いを説明し扱う問題の定義を行った。

第3章では2章で述べた課題に対する解決法として、転移学習の特殊な例である飼いならし学習問題を題材としたデータの利活用を促す手法を提案した。提案手法では、飼いならし学習問題で扱う低信頼かつ多量のデータベースに含まれている所望基準とは異なる基準で与えられたデータに対して与えられるべきラベルを推定・修正し、これから学習したい問題に対して有効に学習可能なデータへと昇華するアルゴリズムを提案した。

第4章では、UCIレポジトリによってまとめられている汎用データセットを用いた提案手法の検証実験を行った。検証実験では定性的なラベルを与えられたデータセットと、定量的なラベルを与えられたデータセットの二種類に対して提案手法を適用した。また、それぞれのデータセットから高信頼な飼育データと低信頼な野生データを無作為に抽出し、さらに野生データからランダムに

選択したデータに対して意図的に誤ったラベルを持たせた疑似データセットを作成し飼いならし学習問題を再現した．検証実験では分類問題を扱い，二値分類問題，多値分類問題に対する検証を行った．検証内容として，提案手法によるラベルの推定精度，提案手法適用前後の野生データの推移の検証，従来手法との精度比較を行った．これらの検証の結果，提案手法を用いることで野生データ中に含まれる誤ったラベルを修正することができ，学習に活用可能なデータを増やせることが示された．また，ラベル修正後のデータを用いて学習することで，飼いならし学習問題の従来手法である BaggingTaming よりも最大で約 10[%] の精度向上を実現した．

第5章では，提案手法の実サンプルに対する実証実験として，ワイヤボンディングの抜取検査における品質推定を実施した．実際に作製したボンディングサンプルを用いて抜取検査を模擬したデータセットを構築し，提案手法を用いた際の道サンプルに対する品質推定能力を検証した．実験の結果，提案手法は BaggingTaming よりも高い品質推定能力を示し，さらに現行の抜取検査で得られるサンプルを用いた機械学習よりも約 20[%] も高い品質推定能力を示した．

以上のように，本研究では誤りを含む多量のデータセットに対して，学習目標に従ったラベル修正による積極的なデータ活用法を提案した．また，汎用データセットと実サンプルを用いた実験から提案手法の有効性を示した．上記の結果から，提案手法はデータ収集・管理コストが高いため高信頼なデータを多く収集できない事例に対して，限られたデータを活用し高い学習性能を示すことができるといえる．

今後の展望について述べる．本論文の冒頭で述べたように，産業分野におけるデータの利活用は世界的に注目を集めており，わが国をはじめとした世界各国において活発に研究・応用されている分野である．したがって，データの利活用に関する国際競争は激化し，あらゆるビジネス・産業において多様化する

ニーズへの対応は急務の課題となる．わが国の製造業の強みである高い現場力や蓄積されたノウハウなどの暗黙知を活用しつつ，オンデマンド生産・多品種少量生産などの多様化したニーズに応答可能な生産構築を進めていき，将来のわが国の産業を育てることでグローバル経済におけるわが国の地位を確立することが重要であると考えている．限られたデータを活用して高い学習性能を示した本研究の成果は，収集コストの高いリアルデータの利活用のハードルを下げ，様々な分野への応用可能性を示唆している．この成果は，第4次産業革命におけるわが国の目指す産業の姿を実現する大きな一歩を踏み出したといえる．

謝辞

本論文は，筆者の九州工業大学大学院生命体工学研究科生命体工学専攻宮本研究室での研究成果をまとめたものです．博士学位論文を提出するにあたって，多くの方々のご指導とご助力をいただきました．

本研究を進めるにあたり，指導教官として御指導，御鞭撻をいただきました九州工業大学大学院生命体工学研究科 宮本弘之 准教授に心から感謝申し上げます．

本論文の審査にあたり，貴重なご指導ご助言を賜りました九州工業大学大学院生命体工学研究科 石井和男 教授，和田親宗 教授，九州工業大学工学研究院機械知能工学研究系 金亨燮 教授に感謝いたします．

国立研究開発法人 産業技術総合研究所 製造技術研究部門 主任研究員 石田秀一氏，産業技術総合研究所 製造技術研究部門 センサシステム技術研究グループ 田原竜夫 グループ長の両名には，筆者のアルゴリズム開発や実験の遂行にあたり，多くの有益なご意見，御指導をいただきました．ここに深い感謝の意を表します．また，本論文の第5章の実験では，産業技術総合研究所

製造技術研究部門 主任研究員 岩崎渉氏に実験に使用する機器，資料を提供していただくとともに有益なご助言をいただきました．ここに感謝の意を表します．

小石 泰毅

関連図書

- [1] 2018 年版ものづくり白書, 2018.
- [2] 中條武志. ヒューマンエラー事例の分類に基づく作業管理システムの評価. 品質, Vol. 23, No. 3, pp. 309–317, 1993.
- [3] 大矢孝, 魚本健人, 堤知明. ニューラルネットワークを用いたコンクリート製造管理システムに関する研究. 土木学会論文集, No. 514, pp. 9–18, 1995.
- [4] 板倉豊和, 服部可奈子, 折原良平. 工場作業者の作業軌跡データからのムダ作業発見. 電子情報通信学会技術研究報告. PRMU, パターン認識・メディア理解, Vol. 106, No. 301, pp. 71–76, 2006.
- [5] 石垣司, 竹中毅, 本村陽一. 日常購買行動に関する大規模データの融合による顧客行動予測システム. 人工知能学会論文誌, Vol. 26, No. 6, pp. 670–681, 2011.
- [6] 本村陽一, 西田佳史, 持丸正明, 赤松幹之, 内藤耕, 橋田浩一. サービスイノベーションのための大規模データの観測・モデリング・サービス設計・適用のループ (< 特集 > サービスイノベーションと ai その 2). 人工知能学会誌, Vol. 23, No. 6, pp. 736–742, 2008.
- [7] 矢田勝俊. スーパーマーケットにおける顧客動線分析と文字列解析. 統計数理, Vol. 56, No. 2, pp. 199–213, 2008.

- [8] 森脇紀彦, 大久保教夫, 早川幹, 佐藤信夫, 福間晋一, 矢野和男, 小野貴司, 妹尾大. 人間行動ビッグデータを活用した店舗業績向上要因の発見. 日本統計学会誌, Vol. 43, No. 1, pp. 69–83, 2013.
- [9] 松村直樹, 和泉潔, 山田健太. Pos データに基づく欠品時の顧客行動を考慮した小売店舗の購買シミュレーション. 人工知能学会論文誌, Vol. 31, No. 2, pp. F–F13_1, 2016.
- [10] 河寄光毅, 吉川大弘, 古橋武. 時系列データの異常検出を目的とした深層学習における再構築誤差の利用可能性に関する検討. 人工知能学会全国大会論文集, Vol. 29, pp. 1–4, 2015.
- [11] 北森俊行. 未経験の異常を予知するためのモデリング. 計測と制御, Vol. 53, No. 2, pp. 98–103, 2014.
- [12] 間野大樹, 是永敦. Ae 及び振動観測による転がり軸受の損傷診断. 精密工学会学術講演会講演論文集 2014 年度精密工学会春季大会, pp. 683–684. 公益社団法人 精密工学会, 2014.
- [13] 北川真也, 木村定雄, 歌川紀之, 森山守. 内部損傷を有するコンクリート床版の打音特性に関する基礎実験. 土木学会論文集 E1 (舗装工学), Vol. 72, No. 1, pp. 1–11, 2016.
- [14] 小野田崇, 伊藤憲彦, 是枝英明. 水力発電所における異常予兆発見支援ツールの開発. 電気学会論文誌 D (産業応用部門誌), Vol. 131, No. 4, pp. 448–457, 2011.
- [15] 小野田崇, 伊藤憲彦, 山崎博信. Svm に基づく水力発電所の異常予兆発見支援. 人工知能学会全国大会論文集 2008 年度人工知能学会全国大会 (第 22 回) 論文集, pp. 218–218. 社団法人 人工知能学会, 2008.

- [16] 高濱正幸, 三上尚高. ガスタービンプラントの異常予兆検知. 品質工学, Vol. 20, No. 4, pp. 45–51, 2012.
- [17] 小野田崇, 村田博士, 山名美智子, 清水健二, 野邊信勝, 村川秀樹. One class svm に基づく水力発電所軸受異常振動の予兆発見 (1). 人工知能学会全国大会論文集 2004 年度人工知能学会全国大会 (第 18 回) 論文集, pp. 170–170. 社団法人 人工知能学会, 2004.
- [18] 渋谷久恵, 前田俊二. 運転パターン情報を利用した異常検知技術. 電気学会論文誌 C (電子・情報・システム部門誌), Vol. 133, No. 10, pp. 1998–2006, 2013.
- [19] 水島晃, 野口伸, 石井一暢, 寺尾日出男. 自律走行車両の gps 位置計測に関わる傾斜補正. 農業機械学会誌, Vol. 62, No. 4, pp. 146–153, 2000.
- [20] 伊藤研一郎, 立山義祐, 西村秀和, 小木哲朗. ヘッドアップディスプレイを用いた自動二輪運転者への情報提示位置の評価. ヒューマンインタフェース学会論文誌 The transactions of Human Interface Society, Vol. 18, No. 1, pp. 435–442, 2016.
- [21] 小野口一則. 自動運転のためのレーン検出アルゴリズム. 電気学会誌, Vol. 135, No. 7, pp. 421–424, 2015.
- [22] 渡邊翔一郎, 佐藤安弘, 古関隆章, 水間毅, 田中龍治, 宮路佳浩, 磯部栄介. 自動列車運転装置を用いた省エネルギー運転制御手法の実証試験. 電気学会論文誌 D (産業応用部門誌), Vol. 137, No. 12, pp. 924–933, 2017.
- [23] 佐藤健哉, 橋本雅文, 菅沼直樹, 加藤真平, 芝直之, 花井將臣, 高田広章, 天沼正行, 沓名守道, 大石淳也. 協調型自動運転のための ldm グローバルコ

- ンセプト実証実験. 第 13 回 ITS シンポジウム講演論文集, pp. 1-1B-10-1 – 1-1B-10-6, 2015.
- [24] Michael Montemerlo, Jan Becker, Suhrid Bhat, Hendrik Dahlkamp, Dmitri Dolgov, Scott Ettinger, Dirk Haehnel, Tim Hilden, Gabe Hoffmann, and Burkhard Huhnke. Junior: The stanford entry in the urban challenge. *Journal of field Robotics*, Vol. 25, No. 9, pp. 569–597, 2008.
- [25] Chris Urmson, Joshua Anhalt, Drew Bagnell, Christopher Baker, Robert Bittner, MN Clark, John Dolan, Dave Duggins, Tugrul Galatali, and Chris Geyer. Autonomous driving in urban environments: Boss and the urban challenge. *Journal of Field Robotics*, Vol. 25, No. 8, pp. 425–466, 2008.
- [26] Julius Ziegler, Henning Latengahn, Markus Schreiber, Christoph G Keller, Carsten Knoppel, Jochen Hipp, Martin Haueis, and Christoph Stiller. Video based localization for berth. In *Intelligent Vehicles Symposium Proceedings, 2014 IEEE*, pp. 1231–1238. IEEE, 2014.
- [27] Jesse Levinson, Jake Askeland, Jan Becker, Jennifer Dolson, David Held, Soeren Kammel, J Zico Kolter, Dirk Langer, Oliver Pink, and Vaughan Pratt. Towards fully autonomous driving: Systems and algorithms. In *Intelligent Vehicles Symposium (IV), 2011 IEEE*, pp. 163–168. IEEE, 2011.
- [28] 菅沼直樹, 魚住剛弘. Gnss/ins と白線検出の融合による自動運転自動車の自己位置推定. 自動車技術会論文集, Vol. 42, No. 5, pp. 1151–1156, 2011.
- [29] 田村浩一郎, 上野山勝也, 飯塚修平, 松尾豊. 深層学習を用いたアンサンブルモデルによる株主価値推定モデルの提案. 人工知能学会論文誌, Vol. 33, No. 1, pp. A-H51_1, 2018.

- [30] 高橋大志. テキスト情報による金融資産価格評価~ データ解析とシミュレーション分析~. 計測と制御, Vol. 55, No. 1, pp. 41–46, 2016.
- [31] 宮崎邦洋, 松尾豊. Deep learning を用いた株価予測の分析. 人工知能学会全国大会論文集 2017 年度人工知能学会全国大会 (第 31 回) 論文集, pp. 2D3OS19a3–2D3OS19a3. 一般社団法人 人工知能学会, 2017.
- [32] 中川慧, 今村光良, 吉田健一. 価格変動パターンを用いた市場予測 k-medoids clustering with indexing dynamic time warping の株式市場への適用. 電気学会論文誌 C (電子・情報・システム部門誌), Vol. 138, No. 8, pp. 986–991, 2018.
- [33] 松井藤五郎, 和泉潔. 新聞記事の時系列テキスト分析による株式市場の動向予測. 人工知能学会全国大会論文集 2016 年度人工知能学会全国大会 (第 30 回) 論文集, pp. 3L3OS16a6–3L3OS16a6. 一般社団法人 人工知能学会, 2016.
- [34] Yue Deng, Feng Bao, Youyong Kong, Zhiquan Ren, and Qionghai Dai. Deep direct reinforcement learning for financial signal representation and trading. *IEEE transactions on neural networks and learning systems*, Vol. 28, No. 3, pp. 653–664, 2017.
- [35] 森井秀幸, 坂本良太, 野村由司彦. 深層学習による麻酔科医の気管挿管手技熟練度判定. IIP 情報・知能・精密機器部門講演会講演論文集 2018, p. 2C10_1. 一般社団法人 日本機械学会, 2018.
- [36] 長谷川晃, 李鎔範, 竹内悠, 市川勝弘. 深層学習を用いた冠動脈 ct における石灰化とステントの自動分類. 日本放射線技術学会雑誌, Vol. 74, No. 10, pp. 1138–1143, 2018.

- [37] 周向榮, 藤田広志. 深層学習に基づく ct 画像からの複数の解剖学的構造の同時自動認識と抽出. *Medical Imaging Technology*, Vol. 35, No. 4, pp. 187–193, 2017.
- [38] 衛藤亮太, 松原靖子, 山下和人, 國澤進, 今中雄一, 櫻井保志. 深層学習を用いた電子カルテ医療情報の多角的解析. 電子情報通信学会技術研究報告, Vol. 117, No. 212, pp. 43–48, 2017.
- [39] 古口眞, 柴田義孝, 橋本浩二. ウェアラブルデバイスを用いた非常時体調管理システム. 第 78 回全国大会講演論文集, Vol. 2016, No. 1, pp. 987–988, 2016.
- [40] 玉城幹介, 桑原恒夫, 山田光一, 中村喜宏, 満永豊, 小西納子, 天野和哉. 個人進度別教育支援システム mesia. 情報処理学会論文誌, Vol. 41, No. 8, pp. 2351–2362, 2000.
- [41] 児玉翠, 秦崇洋, 社家一平. ラーニングレコードを用いた学習者の習熟度時間変化モデルの構築 (教育工学). 電子情報通信学会技術研究報告, Vol. 117, No. 469, pp. 249–252, 2018.
- [42] Qinghua Su. *Potato Shape Grading Using Depth Imaging*. PhD thesis, 2018.
- [43] 平野竜, 高木正則, 佐々木淳, 山田敬三. 圃場データの教材利用に向けた人感センサ検知画像への農作業情報自動付与機能の開発と評価. 第 79 回全国大会講演論文集, Vol. 2017, No. 1, pp. 659–660, 2017.
- [44] 若森和昌, 柴田瞬, 峰野博史. 深層学習を用いた植物の水分ストレス推定手法の検討. マルチメディア, 分散協調とモバイルシンポジウム 2017 論文集, Vol. 2017, pp. 199–206, 2017.

- [45] 三好健悟, 村上幸一. 深層学習を用いた収穫日予測手法の検証. 人工知能学会全国大会論文集 2018 年度人工知能学会全国大会 (第 32 回) 論文集, pp. 2A203–2A203. 一般社団法人 人工知能学会, 2018.
- [46] 相原健郎. ビッグデータを用いた観光動態把握とその活用: 動体データで訪日外客の動きをとらえる. 情報管理, Vol. 59, No. 11, pp. 743–754, 2017.
- [47] 土田崇仁, 遠藤雅樹, 加藤大受, 江原遥, 廣田雅春, 横山昌平, 石川博. Word2vec を用いた地域やランドマークの意味演算. In *DEIM Forum*, pp. H5–1, 2016.
- [48] 高橋雅彦, 土屋陽介, 成田雅彦, 加藤由花. 移動ロボットを利用した案内サービスの構築を支援するプラットフォーム環境. 情報処理学会論文誌, Vol. 55, No. 2, pp. 761–772, 2014.
- [49] Pedro Ballester and Ricardo Matsumura de Araújo. On the performance of googlenet and alexnet applied to sketches. In *AAAI*, pp. 1124–1128, 2016.
- [50] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pp. 1097–1105, 2012.
- [51] 河原達也. 音声認識技術. 電子情報通信学会誌, Vol. 98, No. 8, pp. 710–717, 2015.
- [52] Frank Seide, Gang Li, and Dong Yu. Conversational speech transcription using context-dependent deep neural networks. In *Twelfth annual conference of the international speech communication association*, 2011.

- [53] 加藤和平, 大島考範, 二宮崇. word2vec と深層学習を用いた大規模評判分析. 言語処理学会第 21 回年次大会発表論文集, pp525-528, 2015.
- [54] 小町守, 鈴木久美. 検索ログからの半教師あり意味知識獲得の改善. 人工知能学会論文誌, Vol. 23, No. 3, pp. 217–225, 2008.
- [55] 萩原正人, 関根聡. 半教師あり学習に基づく大規模語彙に対応した日本語単語分割. 言語処理学会第 18 回年次大会発表論文集, pp. 1280–1283, 2012.
- [56] 渡邊陽太郎, 浅原正幸, 松本裕治. 述語語義と意味役割の結合学習のための構造予測モデル. 人工知能学会論文誌, Vol. 25, No. 2, pp. 252–261, 2010.
- [57] 井上裁都, 斎藤博昭. ラベルなしデータの二段階分類とアンサンブル学習に基づく半教師あり日本語語義曖昧性解消. 自然言語処理, Vol. 18, No. 3, pp. 247–271, 2011.
- [58] 立間淳司, 青野雅樹. 多様体ランキングを用いた三次元物体の形状類似検索. 情報処理学会論文誌, Vol. 49, No. 10, pp. 3670–3681, 2008.
- [59] 藤橋一輝, 木村雅之, 金崎朝子, 小澤順. 半教師あり学習による商品画像中の個数と位置の同時推定. 人工知能学会全国大会論文集 2018 年度人工知能学会全国大会 (第 32 回) 論文集, pp. 4M204–4M204. 一般社団法人 人工知能学会, 2018.
- [60] 颯々野学. 日本語単語分割を題材としたサポートベクタマシンの能動学習の実験的研究. 自然言語処理, Vol. 13, No. 2, pp. 27–41, 2006.
- [61] 塚本浩司, 颯々野学. Adaboost と能動学習を用いたテキスト分類. 情報処理学会研究報告自然言語処理 (NL), Vol. 2001, No. 112 (2001-NL-146), pp. 81–89, 2001.

- [62] 山田寛康. Shift-reduce 法に基づく日本語固有表現抽出. 情報処理学会研究報告音声言語情報処理 (SLP), Vol. 2007, No. 47 (2007-SLP-066), pp. 13–18, 2007.
- [63] 森信介, Graham Neubig. 仮名漢字変換ログの活用による言語処理精度の自動向上. 言語処理学会第 16 回年次大会, pp. A1-2-1 – A1-2-4, 2010.
- [64] 浦野貴裕, 酒井薫, 前田俊二, 佐藤真一. 対話的教示欠陥選択に基づく効率的な実欠陥・虚報弁別技術. 電子情報通信学会論文誌 D, Vol. 96, No. 1, pp. 221–229, 2013.
- [65] 三杉大輔, 高橋優介, 笹井一人, 佐藤彰洋, 北形元, 木下哲男. ネットワークトラヒック分析支援エージェントによる能動的異常検知. 情報処理学会研究報告, Vol. 2009, No. 27, pp. 1–6, 2009.
- [66] 小池泰輔, 梅澤猛, 大澤範高. 機械学習型侵入検知システムにおける能動学習を用いた誤検知低減手法の検討. 情報処理学会第 77 回全国大会講演論文集, Vol. 2015, No. 1, pp. 487–488, 2015.
- [67] 野田哲男, 長野陽, 永谷達也, 堂前幸康, 長野鉄明, 田中健一, 小笠原司. 機械学習の枠組みに基づく能動型探索アルゴリズムのサーボパラメータ調整問題への適用性の検討. 計測自動制御学会論文集, Vol. 53, No. 3, pp. 217–228, 2017.
- [68] 前田佳弘, 郷直樹, 岩崎誠. 最適化問題の可解性を利用した制御パラメータの安定範囲算出とパラメータ調整の効率化. 電気学会論文誌 D (産業応用部門誌), Vol. 138, No. 5, pp. 392–401, 2018.

- [69] 馬見塚拓, 安倍直樹. 集団能動学習—データマイニング・バイオインフォマティックスへの展開—. 電子情報通信学会論文誌 D, Vol. 85, No. 5, pp. 717–724, 2002.
- [70] 庄野逸, 鈴木聡志, 木戸尚治. ディープラーニングの医用画像への応用. 医用画像情報学会雑誌, Vol. 33, No. 4, pp. 75–80, 2016.
- [71] 庄野逸, 鈴木藍雅, 鈴木聡志, 木戸尚治. 2 段階転移学習を用いたディープコンボリユーションネットの医用画像認識. 日本神経回路学会誌, Vol. 24, No. 1, pp. 3–12, 2017.
- [72] 則のぞみ, 鹿島久嗣, 山下和人, 猪飼宏, 今中雄一. マルチタスク学習による集中治療室入室患者のリスクモデル構築. 電子情報通信学会論文誌 D, Vol. 100, No. 2, pp. 194–204, 2017.
- [73] 高橋弘樹, 山根順子, 藤渕航. 少ない遺伝子発現データを活用するための畳み込みニューラルネットワークと転移学習による毒性予測システム. 情報処理学会研究報告バイオ情報学 (BIO), Vol. 2018, No. 2, pp. 1–1, 2018.
- [74] 椿真史, 新保仁, 松本裕治. 表現学習と深層学習を用いたタンパク質のフォールド認識と残基接触予測. 人工知能学会研究会, Vol. 5, No. 01, pp. 62–67, 2016.
- [75] Hiromi Oda, Shin Kiyohara, Koji Tsuda, and Teruyasu Mizoguchi. Transfer learning to accelerate interface structure searches. *Journal of the Physical Society of Japan*, Vol. 86, No. 12, p. 123601, 2017.
- [76] Tomohiro Yonezu, Tomoyuki Tamura, Ichiro Takeuchi, and Masayuki Karasuyama. Knowledge-transfer based cost-effective search for interface struc-

- tures: A case study on fcc-al [110] tilt grain boundary. *arXiv preprint arXiv:1708.03130*, pp. 1–26, 2017.
- [77] Maxwell L Hutchinson, Erin Antono, Brenna M Gibbons, Sean Paradiso, Julia Ling, and Bryce Meredig. Overcoming data scarcity with transfer learning. *arXiv preprint arXiv:1711.05099*, pp. 1–10, 2017.
- [78] 戸田隆道, 田中翔太, 井上創造, 上田修功. ラベルの時刻ずれに対応した携帯センサ行動認識手法. マルチメディア, 分散協調とモバイルシンポジウム 2014 論文集, Vol. 2014, pp. 394–400, 2014.
- [79] 中山英樹. 深層畳み込みニューラルネットワークによる画像特徴抽出と転移学習. 信学技報, Vol. 115, No. 146, pp. 55–59, 2015.
- [80] 加納学, 吉崎亮介. ベイズ的最適化と転移学習を用いた製造設備スケールアップ時の操業条件最適化. 計測と制御, Vol. 56, No. 9, pp. 695–698, 2017.
- [81] Michael T Rosenstein, Zvika Marx, Leslie Pack Kaelbling, and Thomas G Dietterich. To transfer or not to transfer. In *NIPS 2005 workshop on transfer learning*, Vol. 898, pp. 1–4, 2005.
- [82] 神寫敏弘, 濱崎雅弘, 赤穂昭太郎. 飼いならしー飼育・野生混在データからの学習. 人工知能学会全国大会 (第 22 回) 論文集, 2D1-3, 2008.
- [83] Leo Breiman. Bagging predictors. *Machine learning*, Vol. 24, No. 2, pp. 123–140, 1996.
- [84] Leo Breiman. Arcing classifier (with discussion and a rejoinder by the author). *The annals of statistics*, Vol. 26, No. 3, pp. 801–849, 1998.

- [85] David Yarowsky. Unsupervised word sense disambiguation rivaling supervised methods. In *Proceedings of the 33rd annual meeting on Association for Computational Linguistics*, pp. 189–196. Association for Computational Linguistics, 1995.
- [86] Avrim Blum and Tom Mitchell. Combining labeled and unlabeled data with co-training. In *Proceedings of the eleventh annual conference on Computational learning theory*, pp. 92–100. ACM, 1998.
- [87] Partha Pratim Talukdar and Fernando Pereira. Experiments in graph-based semi-supervised learning methods for class-instance acquisition. In *Proceedings of the 48th annual meeting of the association for computational linguistics*, pp. 1473–1481. Association for Computational Linguistics, 2010.
- [88] Sinno Jialin Pan and Qiang Yang. A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, Vol. 22, No. 10, pp. 1345–1359, 2010.
- [89] Simon Tong and Daphne Koller. Support vector machine active learning with applications to text classification. *Journal of machine learning research*, Vol. 2, No. Nov, pp. 45–66, 2001.
- [90] Alina Beygelzimer, Sanjoy Dasgupta, and John Langford. Importance weighted active learning. In *Proceedings of the 26th annual international conference on machine learning*, pp. 49–56. ACM, 2009.
- [91] Dana Angluin. Queries and concept learning. *Machine learning*, Vol. 2, No. 4, pp. 319–342, 1988.

- [92] <https://archive.ics.uci.edu/ml/datasets/Wine+Quality>, accessed 2018-6-12.
- [93] <https://archive.ics.uci.edu/ml/datasets/Abalone>, accessed 2018-6-12.
- [94] 塚田裕, 山中公博, 禰占孝之. フリップチップ接続による半導体チップ実装の課題と今後の展望. 電子情報通信学会論文誌 C, Vol. 91, No. 11, pp. 509–518, 2008.
- [95] Michitaka Urushima. Process of mounting tape automated bonded semiconductor chip on printed circuit board through bumps, December 12 1995. US Patent 5,474,957.
- [96] 土師宏, 守田俊章, 中島英治, 吉永日出男. Au-al ワイヤボンディングにおける有機物汚染の影響. 日本金属学会誌, Vol. 56, No. 10, pp. 1127–1131, 1992.
- [97] Shuichi Ishida, Yasutake Koishi, Tatsuo Tabaru, Wataru Iwasaki, and Hiroyuki Miyamoto. Improvement of sampling inspection for wire bonding using taming and thin film ae sensor. *Transactions of The Japan Institute of Electronics Packaging*, Vol. 9, pp. E16–015, 2016.
- [98] Shuichi Ishida, Yasutake Koishi, Tatsuo Tabaru, Wataru Iwasaki, and Hiroyuki Miyamoto. Improvement of sampling inspection for wire bonding using taming and thin film ae sensor. *Transactions of The Japan Institute of Electronics Packaging*, Vol. 9, pp. E16–015, 2016.
- [99] 石坂彰利, 岩田誠一, 山本博司. Au と al の熱圧着における接合界面の変形と新生面の生成. 日本金属学会誌, Vol. 41, No. 11, pp. 1154–1160, 1977.

- [100] 土師宏, 守田俊章, 中島英治, 吉永日出男. ワイヤボンディングにおける超音波振動付加の影響. 日本金属学会誌, Vol. 57, No. 8, pp. 884–889, 1993.
- [101] 里中忍, 上野一也, 平真樹, 後藤進. 超音波を利用した電子部品ワイヤボンディング部の非破壊評価. 溶接学会論文集, Vol. 21, No. 4, pp. 546–552, 2003.
- [102] アコースティック・エミッション試験2 2008. 社団法人日本非破壊検査協会, 2008.

研究業績リスト

誌上発表

国際誌 査読有り 筆頭

1. Yasutake Koishi, Shuichi Ishida, Tatsuo Tabaru, Hiroyuki Miyamoto. Label Estimation Method with Modifications for Unreliable Examples in Taming, International Journal of Networking and Computing, vol.8, No.2, pp. 153–165 (2018)

和文誌 査読有り 筆頭

1. 小石 泰毅, 石田 秀一, 田原 竜夫, 宮本 弘之. 飼いならし学習問題における不採用訓練事例の再活用手法. 電子情報通信学会 D 分冊, J100-D 巻 11 号, pp. 917–925 (2017)
2. 小石 泰毅, 石田 秀一, 田原 竜夫, 岩崎 渉, 宮本 弘之. 薄型 AE センサと転移学習を用いたワイヤボンディング抜取検査の性能向上. エレクトロニクス実装学会誌 (Accepted)

国際誌 査読有り 非筆頭

1. Shuichi Ishida, Yasutake Koishi, Tatsuo Tabaru, Wataru Iwasaki, Hiroyuki Miyamoto. Improvement of Sampling Inspection for Wire Bonding Using

Taming and Thin Film AE Sensor. Transactions of The Japan Institute of Electronics Packaging, Volume 9, pp. E16-015-1–E16-015-7 (2016)

口頭発表・ポスター発表

国際学会 査読有り 筆頭・登壇者

1. Yasutake Koishi, Shuichi Ishida, Tatsuo Tabaru, Hiroyuki Miyamoto. Label Estimation Method with Modifications for Unreliable Examples in Taming. The Fifth International Symposium on Computing and Networking (2017), Aomori (Japan).

国内学会 査読有り 筆頭・登壇者

1. 小石 泰毅, 石田 秀一, 田原竜夫, 宮本弘之. 飼いならし学習問題における信頼性の低い訓練事例に対する真のラベル推定. 第27回日本神経回路学会 (2017), 北九州

国内学会 査読なし 筆頭・登壇者

1. 小石 泰毅, 石田 秀一, 田原竜夫, 宮本弘之. 飼いならし学習問題における誤ラベルデータ検出. 電子情報通信学会ソサイエティ大会 (2016), 札幌
2. 小石 泰毅, 石田 秀一, 田原竜夫, 宮本弘之. 飼いならし学習問題における信頼性の低い訓練事例に対する多値ラベル推定法. 電子情報通信学会ソサイエティ大会 (2017), 東京
3. 小石 泰毅, 宮本 弘之, 石田 秀一, 田原 竜夫. 多様性を考慮した飼いならし学習問題の精度向上に関する研究. 計測自動制御学会九州支部学術講演会 (2017), 鹿児島

4. 小石 泰毅, 石田 秀一, 田原 竜夫, 岩崎 渉, 宮本 弘之. 薄型 AE センサと転移学習を用いたワイヤボンディング抜取検査の性能向上. 第 32 回エレクトロニクス実装学会春季講演大会 (2018), 千葉

国内学会 非筆頭

1. 内田雅人, 小石 泰毅, 宮本弘之, 石田 秀一, 田原竜夫. 飼いならし学習を用いた AE 源の位置標定. 計測自動制御学会九州支部学術講演会 (2017), 鹿児島

受賞・表彰等

[講演大会 優秀賞]

1. 小石 泰毅, 石田 秀一, 田原 竜夫, 岩崎 渉, 宮本 弘之. 薄型 AE センサと転移学習を用いたワイヤボンディング抜取検査の性能向上. 第 32 回エレクトロニクス実装学会春季講演大会 (2018), 千葉

以上